# Molecular mapping of dwarf growth habit trait in apple (*Malus pumila* Mill.) using molecular markers and transcriptomics approaches

by

Zama Thandekile Laureen Mbulawa

Dissertation presented for the degree of
Doctor of Philosophy

at

**Stellenbosch University**
Genetics Department, Faculty of AgriSciences

Supervisor: Dr Aletta E. Bester-van der Merwe
Department of Genetics, Stellenbosch University, South Africa

Co-supervisors: Dr Johan Kriel and Kenneth R. Tobutt
Agricultural Research Council (Infruitec-Nietvoorbij), Stellenbosch, South Africa

March 2020

# Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2020

# Abstract

Apple (*Malus pumila* Mill.) is one of the most important deciduous fruit crops worldwide. Apples are traditionally valued as an important dietary source of fibre and are high in antioxidants, contributing to human nutrition. In South Africa, the apple industry plays a vital role in the country's agricultural economy due to global exports. In recent years, more emphasis has been directed to dwarf trees, as they are well suited for profitable high-density orchards and sustainability of fruit production. However, dwarfism cannot always be linked to increased yield.

At Bien Donné Research Farm of the Agricultural Research Council (ARC) Infruitec-Nietvoorbij', several dwarf growth habits exist, which is related to a form of hybrid incompatibility, hybrid necrosis. One of them is associated with undesirable characteristics such as crinkled leaves and poor growth. Expression of hybrid necrosis in plants can lead to a significant reduction in productivity, due to the deleterious epistatic interactions between alleles that arose from divergent genetic backgrounds. Few, if any, genetic studies have thus far investigated crinkle dwarf growth traits in apple. This study aimed to examine the genetic basis underlying the crinkle dwarf phenotype by employing multidisciplinary approaches that included segregation pattern studies, assessment of self-incompatibility (hybrid incompatibility), molecular mapping and transcriptomic profiling of pooled samples of apical buds and young leaves from normal and from crinkle dwarf phenotypes.

The genetics behind the crinkle dwarf trait was undertaken by studying the segregation patterns of the first filial generation (F1) apple progenies, where parental combinations were heterozygous. Segregation ratios of 9:7 and 3:1 were observed, for which crinkled dwarf phenotypes is expressed when one of the two genes is homozygous recessive (*D-ee* or *ddE*-). Additionally, the involvement of self-incompatibility (*S*) was investigated by identifying the parental *S*-genotypes using PCR based consensus and allele-specific primers of the apple *S*-RNAse gene. Eight parental *S*-genotypes were determined. Herein, the *S*-genotypes of Malling 1 ('M.1') ($S_3S_9$) and TSR1T187 ($S_7S_{24}$) were deduced for the first time.

High-density SNP-based parental genetic linkage maps of ‘McIntosh’ and ‘M.1’ were constructed using the apple 20K Infinium SNP array. The crinkle dwarf trait was mapped on linkage group (LG) 8 in ‘McIntosh’ and on LG2 in ‘M.1’. In the consensus genetic map, crinkle dwarf trait also mapped on LG8. Additionally, the crinkle dwarf trait obtained for the parental genetic maps were validated using Kruskal-Wallis (KW) analysis.

To gain deeper insights into the genes regulating crinkle dwarf phenotype, transcriptome profiles of pooled meristematic tissues of normal and crinkle dwarf phenotypes were generated using RNA-sequencing technology. A total of 921 significantly differentially expressed genes (DEGs), with 763 up-regulated and 158 down-regulated transcripts, were identified. Gene expression analyses revealed that defense signaling and stress-related genes were up-regulated during the expression of crinkle dwarf phenotype along with the activation of several antioxidant proteins/enzymes. The high expression of lactoperoxidase (Class III peroxidase) together with glutathione S-transferase suggests the involvement of reactive oxygen species (ROS). Genes typically encoding for pathogenesis-related proteins (chitinase and pectin), antioxidant enzymes, receptor-like protein (protein serine/threonine phosphatase), as well as alpha-linolenic acid, a precursor of the phytohormone jasmonic acid were all up-regulated during expression of crinkle dwarf phenotype. These findings support the notion that crinkle dwarf phenotype does indeed exhibit hybrid necrosis symptoms. Consequently, an autoimmune response might have been triggered by the allele incompatibilities, in this case between ‘McIntosh’ and ‘M.1’.

Overall, the information generated in this study will aid in designing an in-house screening system for eliminating seedlings carrying crinkle dwarf genes from the ARC breeding material. In future, these findings will also aid in the design of crosses with predictable outcomes and in broadening a sustainable genetic base of the apple cultivars for high productivity orchards, while avoiding raising seedlings with dwarf growth habit associated with crinkled leaves.

# Opsomming

Appel (*Malus pumila* Mill.) is een van die belangrikste sagtevrugtegewasse wêreldwyd. Appels word tradisioneel as 'n belangrike voedingsbron beskou en bevat baie antioksidante wat bydra tot menslike voeding. In Suid-Afrika speel die appelbedryf 'n belangrike rol in die land se landbou-ekonomie as gevolg van wêreldwye uitvoere. In die afgelope paar jaar is meer klem gelê op dwergbome, aangesien dit geskik is vir winsgewende, hoë-digtheid boorde met 'n volhoubare vrugteproduksie. Dwergbome kan egter nie altyd gekoppel word aan verhoogde opbrengste nie.

Op die Bien Donné-navorsingsplaas van die Landbounavorsingsraad (LNR) Infruitec-Nietvoorbij, bestaan verskeie dwerggroeiwyses wat verband hou met 'n vorm van hibriede onverenigbaarheid, hibriede nekrose. Een daarvan hou verband met ongewenste eienskappe soos gekreukelde blare en swak groei. Uitdrukking van hibriede nekrose in plante kan lei tot 'n beduidende afname in produktiwiteit as gevolg van die nadelige epistatiese interaksies tussen allele wat voortspruit uit uiteenlopende genetiese agtergronde. Min, indien enige, genetiese studies het tot dusver ondersoek ingestel na die gekreukelde dwerggroei-eienskappe by appels. Hierdie studie het ten doel gehad om die genetiese basis onderliggend aan die gekreukelde dwergfenotipe te ondersoek deur gebruik te maak van multidissiplinêre benaderings wat insluit die studie van segregasiepatrone, assessering van self-onverenigbaarheid (hibriede onverenigbaarheid), molekulêre kartering en transkriptomiese profilering van saamgestelde monsters van apikale knoppe en jong blare van normale en gekreukelde dwergfenotipes.

Die genetika onderliggend aan die gekreukelde dwerg eienskap is ondersoek deur die segregasiepatrone van die eerste filiale generasie (F1) nageslag te bestudeer, waar ouerkombinasies heterosigoties was. Segregasieverhoudings van 9:7 en 3:1 is waargeneem waar gekreukelde dwergfenotipes uitgedruk word wanneer een van die twee gene homosigoties resessief is (*D-ee* of *ddE*). Verder is die betrokkenheid van self-onverenigbaarheid (*S*) ondersoek deur die ouerlike *S*-genotipes, met behulp van PKR-gebaseerde konsensus en alleelspesifieke inleiers van die appel *S*-RNAse geen, te identifiseer. Agt ouerlike *S*-genotipes is bepaal. Die *S*-genotipes van Malling 1 ('M.1') ($S_3S_9$) en TSR1T187 ($S_7S_{24}$) is vir die eerste keer in hierdie studie bepaal.

Hoë-digtheid SNP-gebaseerde ouer genetiese koppelingskaarte van 'McIntosh' en 'M.1' is saamgestel met behulp van die appel 20K Infinium “SNP-array”. Die gekreukelde dwerg eienskap is gekarteer op die koppelingsgroep (“linkage group”, LG) 8 in 'McIntosh' en op LG2 in 'M.1'. Op die konsensus genetiese kaart word gekreukelde dwerg eienskap ook op LG8 gekarteer. Die gekreukelde dwerg eienskap wat verkry is vir die ouerlike genetiese kaarte is ook bevestig met behulp van Kruskal-Wallis (KW) analise.

Om dieper insigte te verkry in die gene wat die gekreukelde dwergfenotipe reguleer, is transkriptoom profiele van saamgevoegde meristematiese weefsels van normale en gekreukelde dwergfenotipes gegenereer met behulp van RNA-opeenvolgingstegnologie. 'n Totaal van 921 gene wat betekenisvol differensieel uitgedruk word (“differentially expressed genes”, DEG's), met 763 op-gereguleerde en 158 af-gereguleerde transkripte, is geïdentifiseer. Hierdie geen uitdrukkings analises het aan die lig gebring dat die verdedigings en stresverwante gene tydens die uitdrukking van 'n gekreukelde dwergfenotipe op-gereguleer is, tesame met die aktivering van verskeie antioksidantproteïene / ensieme. Die hoë uitdrukking van laktoperoksidase (Klas III peroksidase), tesame met “glutathion S-transferase” dui op die betrokkenheid van reaktiewe suurstofspesies (“reactive oxygen species “, ROS). Gene wat vir patogeenverwante proteïene (chitinase en pektien), antioksidantensieme, reseptoragtige proteïene (proteïne serien / treonien fosfatase) kodeer, asook alfa-linoleensuur, 'n voorloper van die fitohormoon jasmonsuur, is almal tydens die uitdrukking van die gekreukelde dwergfenotipe op-gereguleer. Hierdie bevindings ondersteun die gedagte dat die gekreukelde dwergfenotipe wel hibriede nekrose simptome vertoon. 'n Outo-immuunrespons kon gevolglike veroorsaak geword het deur alleelonverenigbaarheid, in hierdie geval tussen 'McIntosh' en 'M.1'.

Gevolglik sal die inligting wat in hierdie studie gegenereer is, help om 'n interne siftingstelsel daar te stel om saailinge wat gekreukelde dwerggene het, te elimineer uit die LNR-teelmateriaal. In die toekoms sal hierdie bevindinge ook help met die samestelling van kruisings met voorspelbare uitkomste en die daarstel van 'n volhoubare genetiese basis van appelkultivars, vir boorde met 'n hoë produktiwiteit, terwyl die dwerggroeiwyse verbonde aan gekreukelde blare vermy word.

# Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions (in no particular order):

Dr Aletta Bester-van der Merwe, for her very valuable inputs, advice and encouragement during the course of my study. Thank you for your great concern, patience and kindness. Thank you for pushing me the hardest, “as hard as it gets” the experience has taught me that, anything is possible. Thank you for making it possible.

My co-supervisor Kenneth R. Tobutt. Your unconditional support as a supervisor has been an inspiration throughout my studies. Thank you for shaping me to be a true Geneticist. I am deeply grateful and I will be forever indebted.

Dr Michela Troggio and Dr Alessandro Cestaro (FEM, Italy) for inviting me to pursue my molecular work and transcriptomics at FEM. Thank you for all the molecular and bioinformatics skills instilled in me. Thank you immensely for sharing your scientific wisdom. Thank you for making Chapters 4 and 5 possible.

Dr Khashief Soeker, you have been a pillar of strength. You went beyond the scope of your work to be my “in-formal supervisor”. You stood by me through thick and thin making sure my studies are a success.

Dr Cecilia Bester (ARC Infruitec) for being my mentor and administrator throughout my PhD project. Thank you for all your guidance, inputs and encouragements. Thank you also for creating opportunities to attend international conferences and for my three-months study visit at FEM, Italy. You have been the calm in the middle of the greatest storm.

Charlie Horgarth and his team at the ARC Infruitec Bien Donne Research Farm, thank you for all the help from generating crosses and taking care of my mapping populations.

My ARC family (Dr Yolanda Petersen, Dr Khashief Soeker, Sieyaam Safodien and Wendy Langerhoven) who afforded me immense support and courage through good and toughest times.

Dr Beatrix Coetzee. Thank you for your great advice pertaining to transcriptomics work and for reviewing my chapters. Above all, thank you for standing in on my last lap when all was a great challenge.

My family who had to endure my absence during the tenure of my doctoral candidature. I love you and appreciate everything you have done for me. “we may grow apart in different directions like branches on a tree yet our roots remain the same”.

My PhD studies were funded by ARC Infruitec-Nietvoorbij through the Technology and Human Resources for Industry Program (THRIP) and the National Research Foundation (NRF) of South Africa. Thank you for the financial support.

Glory be upon the Almighty God “Isaiah 60:22”

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| 2-ODD | 2-Oxoglutarate dependent dioxygenase |
| 3'-end | Three prime end |
| 5'-end | Five prime end |
| ABA | Abscisic acid |
| AFLP | Amplified Fragment Length Polymorphism |
| ARC | Agricultural Research Council |
| BLAST | Basic local alignment search tool |
| BR | Brassinosteroid |
| cDNA | Complementary DNA |
| COG | Cluster of orthologous groups of proteins |
| DEG | Differentially expressed genes |
| DNA | deoxyribonucleic acid |
| DH | Doubled haploid |
| dNTP | Deoxyribonucleotide Triphosphate |
| e-value | expect value |
| EC | Enzyme commission |
| F | Forward primer |
| FDR | False discovery rate |
| FPKM | Fragments Per Kilobase of transcript per Million mapped reads |

| | |
|---|---|
| F1 | First filial generation |
| F2 | Sec filial generation |
| GA | Gibberellin |
| GAI | Gibberellin insensitive |
| GA2ox | Gibberellin 2-beta-dioxygenase |
| GA3ox | Gibberellin 3-beta-dioxygenase |
| GA20ox | Gibberellin 20-beta-dioxygenase |
| gDNA | Genomic deoxyribonucleic acid |
| GID | Gibberellin insensitive dwarf |
| GO | Gene ontology |
| IAA | Indole-3-acetic acid |
| IM | Interval Mapping |
| KAO | *ent*-kaurenoic acid oxidase |
| KEGG | Kyoto encyclopedia of genes and genomes |
| KO | *ent*-kaurene oxidase |
| KS | *ent*-kaurene synthase |
| KW | Kruskal-Wallis |
| LD | Linkage disequilibrium |
| LG | Linkage group |
| LOD | Logarithm of the odds |
| MAS | Marker assisted selection |
| NCBI | National Center for Biotechnology Information |

| | |
|---|---|
| NGS | Next-generation sequencing |
| PCR | Polymerase Chain Reaction |
| pH | Potential of hydrogen |
| p-value | Probability value (as a statistically significant limit) |
| Q | Phred quality score |
| qPCR | Quantitative real-time PCR |
| QTL | Quantitative Trait Loci |
| R | Reverse |
| RAPD | Random Amplified Polymorphic DNA |
| REC | Recombination frequency |
| RFLP | Restriction Fragment Length Polymorphism |
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| RNA-seq | Ribonucleic acid sequencing |
| RPKM | Reads Per Kilobase of transcript per Million mapped reads |
| RT | Room temperature |
| *S* (locus) | Self-incompatibility (locus) |
| SNP | Single nucleotide polymorphism |
| SSR | Simple Sequence Repeat |
| OD | Optical density |
| UK | United Kingdom |
| UV | Ultra Violet |

| | |
|---|---|
| v | Version |

## List of Chemicals

| | |
|---|---|
| CTAB | Cetyltrimethylammonium bromide |
| $dH_2O$ | Distilled water |
| EDTA | Ethylenediaminetetraacetic acid |
| NaCl | Sodium chloride |
| NaOH | Sodium hydroxide |
| PVP-40 | Polyvinylpyrrolidone, average molecular weight 40,000 |
| TAE | Tris-acetate-EDTA |
| TBE | Tris-borate-EDTA |
| TE | Tris HCl-EDTA |
| Tris-HCl | Tris-hydrochloride |
| Tris | Tris-hydroxymethyl-aminomethane |
| TE | Tris EDTA |

## List of Units

| | |
|---|---|
| % | Percentage |
| < | Less than |
| > | Greater than |
| bp | basepairs |
| °C | degrees Celsius |
| cM | centiMorgan |
| g | gram |

| | |
|---|---|
| g/L | gram per litre |
| Gb | Gigabase(s) |
| hr | hour |
| Kb | Kilobase pairs |
| M | Molar |
| Mb | millionbases |
| mg/mL | milligram per milliliter |
| mM | milliMolar |
| min | minutes |
| µg | microgram |
| µL | microliter |
| µM | micromolar |
| ng | nanogram |
| rpm | Revolutions per minute |
| v/v | Volume per volume |
| w/v | Weight per volume |

# Chapter 1
# General Introduction

## 1.1 Research Context and Rationale

Apple (*Malus pumila* Mill.) is one of the earliest fruits utilized by humans, and its domestication dates back at least 3 800 years to the ancient Greeks and Romans (Janick *et al.*, 1996; Harris *et al.*, 2002; Forsline *et al.*, 2003). In South Africa, apples were first planted in the 1650s in Western Cape Province; however, the first commercial apple orchards only started in the late 19$^{th}$ and early 20$^{th}$ century (Hancock *et al.*, 2008).

South Africa's main apple producing areas are in the Western Cape Province and Langkloof East in the Eastern Cape Province, and there are other, smaller, production areas elsewhere in the country (DAFF, 2018; HORTGRO, 2018; Sikuka, 2019). The Western Cape Province, with its Mediterranean-like climate favourable for apple production, accounts for more than half of all the apples produced in South Africa (Ntshidi *et al.*, 2018; Phaleng and Tshitiza, 2018; Sikuka, 2019). South Africa ranks 16$^{th}$ in terms of world production, placing it the largest producer in Africa, with a projected 840 000 metric ton in the 2018-2019 season after recovering from the drought impact experienced in 2016-2017 (Phaleng and Tshitiza, 2018; Sikuka, 2019; USDA-FAS, 2019). Apples account for about 28% of the total deciduous cultivated fruit area in South Africa, with 24 176 hectares in 2017-2018 season (HORTGRO, 2018).

The South African apple industry is export oriented and currently South Africa is the 7$^{th}$ global leading exporter of apples, with 393 344 metric ton exported in 2018 to the Far East and Asia (31%), followed by the African market (30%), the UK (18%), the Middle East (7%) and Europe (6%) (HORTGRO, 2018; Ntshidi *et al.*, 2018). These exports accounts for about 80% of the apple industry's income and contributes significantly to the Western Cape gross domestic product (DAFF, 2018; HORTGRO, 2018; Kuschke and Cassim, 2019). In South Africa, the biggest apple breeding programme is conducted by the Agricultural Research

Council (ARC) in the Crop Development (Breeding and Evaluation) Division of ARC Infruitec-Nietvoorbij, based in Stellenbosch.

There has been an increasing interest in the genetics of dwarf growth habits in apple. The development of dwarf cultivars of crop plants has played a significant role in agriculture during the advent of the "Green Revolution" phenomenon (Peng *et al.*, 1999; Khush, 2001; Sasaki *et al.*, 2002; Hedden, 2003). An ideal dwarf plant type has been proposed as being short in stature, early maturing, with a high harvest index and rapid growth rate (Barthélémy and Caraglio, 2007; Byrne, 2012; Hollender and Dardick, 2015).

Alston (1976) reported three apparently recessive forms of dwarf type in apple: early, crinkle and sturdy dwarfs. One, crinkle dwarf, appears similar to a phenotype segregating in some progenies of 'McIntosh' growing at the ARC Infruitec-Nietvoorbij's Bien Donné Research Farm. A typical dwarf seedling phenotype is associated with crinkled leaves, poor growth, and in some cases lethality. In particular, the crinkle dwarf trait affects plants at the seedling stage, making them unsuitable for production. It is considered to be directly related to hybrid incompatibility in the form of hybrid necrosis and are therefore often associated with agronomically undesirable traits. This deleterious trait is not well-characterised genetically and has not been mapped. Therefore, understanding the genes involved in controlling these phenotypes is important. Identifying markers linked to the trait is of interest in avoiding "carriers" in the ARC and other breeding programmes.

Understanding the genetic basis of important traits requires both phenotypic and genotypic datasets. Fortunately, apple is one of the well-characterised fruit tree species. Progress in apple genomics was revolutionized when the apple genome was first sequenced in 2010 (Velasco *et al.*, 2010), and has since advanced remarkably, with the release of the high quality 'Golden Delicious' double haploid reference genome, GDDH13 v1.1 in 2017 (Daccord *et al.*, 2017). Moreover, the application of next-generation sequencing (NGS) technologies has led to development of the high-throughput single nucleotide polymorphism (SNP) genotyping arrays, the International RosBREED SNP Consortium (IRSC) apple 8K, the FruitBreedomics 20K Illumina Infinium® and the Axiom® 480K SNP arrays, which enable genome-wide screening (Chagné *et al.*, 2012; Bianco *et al.*, 2014; 2016). Additionally,

the advent of RNA sequencing and bioinformatics tools enables in-depth transcriptome profiling of phenotypes between different conditions so that a change in the expression of genes can be identified and eventually quantified (Morozova *et al.*, 2009; Wang *et al.*, 2009; Conesa *et al.*, 2016).

## 1.2 Research Aims and Objectives

This study aims to characterize the crinkle dwarf phenotype and to investigate its inheritance, by combining classical molecular mapping and more advanced transcriptomic approaches.

The specific objectives of the study were:

- to investigate the inheritance patterns of crinkle dwarf phenotype in various mapping populations,
- to identify molecular markers linked to the crinkle dwarf trait in segregating progenies by genotyping with SNP markers followed by co-segregation analysis and investigation of the relevant region of the published apple genome to identify potential candidate genes,
- to identify genes differentially expressed between normal *versus* crinkle dwarf phenotypes using RNA-Seq transcriptome analysis based on the Illumina high-throughput platform.

## 1.3 Chapter Layout

This dissertation is divided into six chapters, including the introduction, the review of literature, three research chapters and the overall conclusion. Each chapter is individually introduced and includes its relevant reference list. The supplementary tables are provided as appendices at the end of the thesis.

**Chapter 1: General Introduction**

A general introduction presents the context pertaining to the importance of the apple industry in South Africa and its significance as well as the background of the research leading to the

rationale of the study. It then briefly explain the aims and objectives, with summary of the chapter layouts.

**Chapter 2: Review of Literature**

A detailed review of literature relating to apple origin and its context is given, including biological and horticultural aspects in terms of botanical features including self-incompatibility and hybrid incompatibility. This chapter also describes an overview of molecular markers and transcriptomics resources available.

**Chapter 3: Genetic basis on the mode of inheritance and *S*-linkage underlying crinkle dwarf growth habit trait in apple through investigating phenotypic segregation patterns**

Clarification on the mode of inheritance underlying dwarf seedlings associated with crinkle leaves is presented, through investigating phenotypic segregation patterns and determination of *S*-linkage in various apple mapping populations.

**Chapter 4: Molecular characterisation and mapping of genes associated with crinkle dwarf trait in apple**

Construction of high density SNP-based genetic linkage maps with the view to mapping genes associated with the crinkle dwarf trait in a progeny from the cross of 'McIntosh' x 'M.1' using the apple 20 K Infinium SNP array are described.

A manuscript entitled "Molecular characterization and mapping of crinkle dwarf trait in apple" is in preparation, to be submitted to a peer-reviewed journal (Molecular Breeding Journal).

**Chapter 5: Differential gene expression between the normal and crinkle dwarf phenotypes in apple based on RNA-Seq transcriptome analysis**

The transcriptomes of pooled samples of apical buds and young leaves from normal and from crinkle dwarf phenotypes sampled at different developmental stages were analysed using the

next-generation sequencing technology, RNA-seq based on the Illumina high-throughput platform. The significance of the resultant differential expressed genes between the contrasting phenotypes is discussed.

**Chapter 6: General Discussions and Future Prospects**

This chapter presents a concluding summary of the main findings. Furthermore, it outlines the limitations of the study, and investigates opportunities for further research.

## 1.4 References

**Alston FH**. 1976. Dwarfing and lethal genes in apple progenies. Euphytica **25**, 505–514.

**Barthélémy D, Caraglio Y**. 2007. Plant architecture: A dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. Annals of Botany **99**, 375–407.

**Bianco L, Cestaro A, Linsmith G, *et al.*** 2016. Development and validation of the Axiom® apple 480K SNP genotyping array. Plant Journal **86**, 62–74.

**Bianco L, Cestaro A, Sargent DJ, *et al.*** 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh.). PLoS One **9**, e110377.

**Byrne DH**. 2012. Trends in fruit breeding. *In*: Badenes ML and Byrne DH, (Eds). *Fruit Breeding*. Boston, MA: Springer US, pp. 3–36.

**Chagné D, Crowhurst RN, Troggio M, *et al.*** 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One **7**, e31745.

**Conesa A, Madrigal P, Tarazona S, *et al.*** 2016. A survey of best practices for RNA-seq data analysis. Genome Biology **17**, 1–19.

**Daccord N, Celton JM, Linsmith G, *et al.*** 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics **49**,

1099–1106.

**DAFF**. 2018. A profile of the South African apple market value chain. Pretoria: Directorate Marketing. South African Department of Agriculture Fisheries and Forestry. http://www.daff.gov.za.

**Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC**. 2003. Collection, maintenance, characterization, and utilization of wild apples of Central Asia. *In*: Janick J, (Eds). *Horticultural Reviews*. John Wiley & Sons, Inc. Vol 29.1–62.

**Hancock JF, Luby JJ, Brown SK, Lobos GA**. 2008. Apples. *In*: Hancock JF, (Eds). *Temperate Fruit Crop Breeding*. Springer, Netherlands. pp. 1-38.

**Harris SA, Robinson JP, Juniper BE**. 2002. Genetic clues to the origin of the apple. Trends in Genetics **18**, 426–430.

**Hedden P**. 2003. The genes of the Green Revolution. Trends in Genetics **19**, 5–9.

**Hollender CA, Dardick C**. 2015. Molecular basis of angiosperm tree architecture. New Phytologist **206**, 541–556.

**HORTGRO**. 2018. Key Deciduous Fruit Statistics: Pome Fruit. http://www.hortgro.co.za.

**Janick J, Cummins J, Brown S, Hemmat M**. 1996. Apples. *In*: Janick J and Moore JN, (Eds). *Fruit Breeding, Tree and Tropical Fruits*. John Wiley & Sons, Inc., Vol I.1–77.

**Khush GS**. 2001. Green revolution: The way forward. Nature Reviews Genetics **2**, 815–822.

**Kuschke I, Cassim A**. 2019. Sustainable agriculture: Market intelligence report. GreenCape.

**Morozova O, Hirst M, Marra M**. 2009. Applications of new sequencing technologies for transcriptome analysis. Annual Review of Genomics and Human Genetics **10**, 135–151.

**Ntshidi Z, Dzikiti S, Mazvimavi D**. 2018. Water use dynamics of young and mature apple trees planted in South African orchards: A case study of the Golden Delicious and Cripps' Pink cultivars. Proceedings of the International Association of Hydrological Sciences **378**,

79–83.

**Peng J, Richards DE, Hartley NM, *et al.*** 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature **400**, 256–261.

**Phaleng L, Tshitiza O**. 2018. South African Fruit Trade Flow. National Agricultural Marketing Council. http://www.namc.co.za.

**Sasaki A, Ashikari M, Ueguchi-Tanaka M, *et al.*** 2002. A mutant gibberellin-synthesis gene in rice. Nature **416**, 701–702.

**Sikuka W**. 2019. South African deciduous fruit exports continue positive growth. USDA Foreign Agricultural Service. Global Agricultural Information Network. Gain Report SA1914.

**USDA-FAS**. 2019. Fresh apples, grapes, and pears: World market and trade. United States Department of Agriculture - Foreign Agricultural Service. Washington, DC: USDA. June.

**Velasco R, Zharkikh A, Affourtit J, *et al.*** 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genetics **42**, 833–839.

**Wang Z, Gerstein M, Snyder M**. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics **10**, 57–63.

## Online resources

Department of Agriculture, Forestry and Fisheries (DAFF): http://www.daff.gov.za

HORTGRO: http://www.hortgro.co.za

GreenCape: https://www.greencape.co.za

National Agricultural Marketing Council (NAMC): http://www.namc.co.za

USDA Foreign Agricultural Service (USDA-FAS): https://www.fas.usda.gov/

# Chapter 2
# Review of Literature

## 2.1 Introduction

More than 77.2 million metric ton of apples are produced worldwide annually, with China being the largest producer and accounting for about 41.4 million metric ton in the year 2017 of the global world production, followed by the United States, Turkey, Poland, Italy, India and France (FAOSTAT, 2018). In the year 2017, South Africa produced about 924 375 metric ton of apples per annum (HORTGRO, 2018), and is the largest producer of apples in Africa, ranking 16$^{th}$ worldwide (DAFF, 2018; FAOSTAT, 2018).

In South Africa, apples were first planted in the 1650s near Cape Town (Western Cape Province) to sustain settlers and supply ships of the Dutch East India Company. The first commercial apple orchards also started in the Western Cape Province in the late 19$^{th}$ and early 20$^{th}$ century especially to supply the export markets opened up by the development of refrigerated shipping. Apple production replaced the faltering wine industry (Hancock *et al.*, 2008). Currently, the main apple producing areas in South Africa are the Western Cape Province (Ceres, Elgin, Grabouw, Groenland, Villiersdorp and Vyeboom) and Eastern Cape Province (Langkloof East) (Figure 2.1). The Western Cape Province accounts for more than half of all the apples produced in South Africa due to its Mediterranean-like climate, which is favourable for apple production (DAFF, 2018; HORTGRO, 2018).


**Figure 2.1 Apple production areas in South Africa.** The red dots denotes the main apple production areas and the black dots represents other small apple production areas. (Adapted from http://www.delecta.co.za/pome-fruit/).

Apples are one of the most economically important deciduous fruits grown in South Africa. The apple industry plays a vital role in South African horticulture, representing approximately 28% of the total deciduous cultivated fruit area, sec only to grape. In addition, about 37% of apples produced locally are destined towards the export market, especially Europe, representing 38% of deciduous fruit exports. This accounts for about 80% of the apple industry's income and contributes greatly to the Western Cape Province gross domestic product (DAFF, 2018; HORTGRO, 2018).

However, the drought in the Eastern Cape and Western Cape Provinces caused by the below average winter rainfall received in 2016-2017 and the low dam levels which has severely

impacted the availability of irrigation water resulted in lower apple yields and smaller fruit sizes over the last few years. Consequently, the 2017-2018 apple production was forecast to decrease to about 800 000 metric ton per annum. Similarly, the exports were also forecast to decrease to about 485 000 metric ton per annum, based on the available production and on fruit not meeting the export quality supplies standards (HORTGRO, 2018; Sikuka, 2019; USDA-FAS, 2019).

## 2.2 Origin, Distribution and Taxonomy of domesticated apple

Based on archaeological and historical evidence, apple represents a relatively recent addition to the list of domesticated plants. The first references to the apple are found in works by Homer, Alexander the Great and in the Bible (Morgan and Richards, 1993). From the human point of view, the beginning of apple cultivation dates back to at least 3 800 years before present (BP) to the ancient Greeks and Romans, coinciding with the records of human use of apples and grafting techniques for asexual tree propagation (Forsline *et al.*, 2003; Cornille *et al.*, 2013). These records suggest that apples have been used by prehistoric humans (Janick *et al.*, 1996; Harris *et al.*, 2002).

The spread of apple from Asia followed along the ancient Silk Road to Europe through human travels and invasions in which horses and donkeys likely carried apple seeds into their guts and unwittingly spread them along the trade routes (Janick *et al.*, 1996; Robinson *et al.*, 2001; Cornille *et al.*, 2012). Harris *et al*. (2002) reported that, during this transfer path of domestication, introgressions into the genome of domesticated apples must have occurred from other wild *Malus* species. Since then, apple has been distributed into almost all parts of the world and this has allowed adapted types to be selected for different environments and regions (Janick *et al.*, 1996; Harris *et al.*, 2002).

### 2.2.1 Taxonomy and nomenclature

The domesticated apple (*Malus pumila* Mill.) belongs to the genus *Malus* of the *Rosaceae* family, under the tribe *Pyreae*. Previously, it was placed within the subfamily *Maloideae* (formerly *Pomoideae*), but a later molecular phylogenetic revision based on the analysis of

sequences from multiple chloroplast and nuclear genes places it within the reformulated *Spiraeoideae* in the clade/subtribe, *Pyrinae* (Potter *et al.*, 2002, 2007). Related deciduous fruit crops include European pear (*Pyrus communis* L.), loquat (*Eriobotrya japonica* Lindl.) and medlar (*Mespilus germanica*), all of the former *Maloideae*, and peach (*Prunus persica* L.), almond (*Prunus dulcis* Mill.) and strawberry (*Fragaria* x *ananassa* L.), as well as ornamental plants including rose (*Rosa* spp.) (Potter *et al.*, 2002, 2007; Shulaev *et al.*, 2008).

Until today there is no clear consensus amongst taxonomists and botanists on the number of species in the genus *Malus*. Most authorities and later reports range from 25-47 (Robinson *et al.*, 2001) up to 55 (Harris *et al.*, 2002) and several subspecies of so-called "crabs" apple which are mostly used as ornamental trees for their attractive flowers and fruits (Janick *et al.*, 1996; Campbell *et al.*, 2007). Apple has an autopolyploidy origin, but through years of domestication its genome has become diploidised (Velasco *et al.*, 2010; Han *et al.*, 2011). The majority of apple cultivars are diploid with $2n=2x=34$, with the exception of some triploid cultivars such as 'Jonagold' and 'Mutsu' (Janick *et al.*, 1996). Some species show variable levels of ploidy (Pereira-Lorenzo *et al.*, 2009).

There has been a lot of dispute on the correct scientific name of the domesticated apple. In terms of morphological and molecular datasets (nuclear ribosomal DNA (rDNA) and chloroplast DNA (cpDNA), a closer relatedness was found between the domesticated apple and the wild apple *M. sieversii*. Therefore, it was proposed that *Malus sieversii* would be the wild ancestor of the domesticated apple, but it is not a distinct species (Robinson *et al.*, 2001; Harris *et al.*, 2002; Forsline *et al.*, 2003; Volk *et al.*, 2009). Other *Malus* species; *M. baccata*, *M. mandshurica*, *M. orientalis*, *M. prunifolia*, *M. sylvestris* are also believed to have contributed to the apple genetic pool although their degree of parentage is poorly understood (Harris *et al.*, 2002). Together with *M. sieversii*, the domesticated apple was referred to as *M. pumila* Mill (Mabberley *et al.*, 2001; Robinson *et al.*, 2001). However, the name *M. pumila*, originally applied to the 'Paradise' rootstocks, and has priority over the most utilised name *Malus* × *domestica* Borkhausen (Borkh.), the name referring to its supposed interspecific origin (Korban and Skirvin, 1984). The completion of the apple genome, cultivar 'Golden Delicious' has provided some insight on the phylogenetic origin of *Malus* phylogenetics, and

it supports the debate on the exact scientific name of the domesticated apple that both are the same species (Velasco *et al.*, 2010). This study uses the scientific name, *Malus pumila*.

With regards to geographical origin, the domesticated apple remains unknown but it is believed to have originated in the mountainous areas of southeast Kazakhstan, Kyrgyzstan, and Tajikistan in Central Asia, where the greatest diversity of wild *Malus* species germplasm is still found (Janick *et al.*, 1996; Harris *et al.*, 2002).

## 2.3 Botany of the domestic apple

### 2.3.1 Apple morphology, pollination and fertilisation

The domestic apple is an outbreeding, deciduous, perennial woody plant (Janick *et al*., 1996). Morphologically, apple tree is small to medium size. The flowers are pentamerous with creamy white to pink petals and having an expanded inferior ovary surrounded by the cortex. The apple fruit is a "pome", which is a fleshy accessory fruit consisting of a central core with seeds enclosed by a papery capsule of fused carpels. The centre of the fruit contains about five carpels arranged in a five-pointed star shape (Janick *et al.*, 1996; Pereira-Lorenzo *et al.*, 2009).

In most angiosperms, flowers must be pollinated to set fruits and, hence, pollination is the primary step in the plant's reproductive success (Stephenson, 1981). During the pollination process, the stigma receives pollen which fertilises ovules to elicit fruit set (Cane and Schiffhauer, 2003). Apple trees are monoecious with both sexes within the same flower and are self-incompatible, showing moderate to severe inbreeding depression (Janick *et al*., 1996; Pereria-Lorenzo *et al*., 2009). They generally need to be cross-pollinated with compatible pollen in order to set fruit (Pereria-Lorenzo *et al*., 2009) and are mostly pollinated by bees. However, some apple cultivars are partially self-compatible (detailed in section 2.3.2).

Apple seeds are dormant at harvest; they cannot germinate directly after being extracted from a fruit (Dennis, 1994) even if all growth conditions are favourable. Apple seeds require a period of cold treatment at 0-5 ºC for about 12 weeks before they can germinate, known as

stratification (Lewak, 2011). At the end of stratification period, germinated seeds may be sown in trays and allowed to grow into seedlings in the glasshouse. Once the first true leaves appear, seedling evaluations may begin.

### 2.3.2 Self-incompatibility in apple

Self-incompatibility was first defined by Darwin (1876), as cited by McClure (2009), as the inability of fertile plants to reproduce after selfing. de Nettancourt (1977) came up with a new definition where he referred to self-incompatibility as the inability of a fertile monoecious seed plant to produce zygotes after self-pollination. A number of underlying genetic mechanisms have been described, with the most common being the *S*-RNase-mediated gametophytic self-incompatibility (GSI) system. The latter is exhibited in *Rosaceae* and in two other distantly-related families: the *Solanaceae* and the *Plantaginacea* (McClure *et al.*, 1989; Sassa *et al.*, 1992). The GSI system is a reproductive barrier that is characterised as a stylar reaction where the arrest of incompatible pollen tubes occur in the style during pollen-tube growth; thus preventing the delivery of sperm cells to the ovary (Kao and Tsukamoto, 2004).

In apple, self-incompatibility is of a gametophytic type and controlled by a single multi-allelic locus, the *S*-locus, derived from the word "self-sterility" (Kobel, 1939; Bateman, 1955; Broothaerts, 2003). This *S*-locus consists of at least two linked gene: *S*-RNase, which acts as a determinant recognition specificity of a pistil (female) (Anderson *et al.*, 1986; Xue *et al.*, 1996), and *S*-haplotype-specific F-box gene (SFB) or *S*-locus F-box (SLF) which acts as determinant candidates for recognition specificity of pollen (male) (Ushijima *et al.*, 2003; Kao and Tsukamoto, 2004). These two *S*-genes, *S*-RNase and SFB/SLB, are closely linked and variants of the *S*-locus are referred to as *S*-haplotypes while the term allele is used to denote variants of one of the given polymorphic genes (Kao and McCubbin, 1996; Takayama and Isogai, 2005).

In the GSI system, when a pollen *S*-allele matches either one of the *S*-alleles of the pistil, *S*-RNase secreted by the pistil tissue degrades the ribosomal RNA produced by the pollen grains, resulting in the inhibition of pollen tube elongation, thus preventing fertilization

(Gibbs, 1988; Kao and Huang, 1994; McClure and Franklin-Tong, 2006). Identification of the *S*-genotypes is thus essential for selecting suitable pollen donors, especially for commercial fruit production and breeding programs (Matsumoto, 2014; Okada, 2015). The success of the GSI mechanism is dependent on the combination of the *S*-haplotypes, as illustrated in Figure 2.2 (Sassa *et al.*, 2007; Kim *et al.*, 2016; Matsumoto and Tao, 2016). The parental cross of $S_1S_2$ x $S_1S_2$ is rendered incompatible when the *S*-alleles of both the pistil and pollen matches, while the cross of $S_1S_2$ x $S_1S_3$ results in semi-compatibility and is mainly governed by the direction of the cross. On the other hand, the parental cross of $S_1S_2$ x $S_3S_4$ has no *S*-allele match and results in full compatibility. The degree of self-incompatibility varies in triploid cultivars, resulting in fruit set when pollinated with diploids, but fluctuates considerably when crossed with other triploids. Bošković and Tobutt (1996) reported that some apple cultivars are cross incompatible, and some triploid x diploid combinations fail whereas the reciprocals succeed.

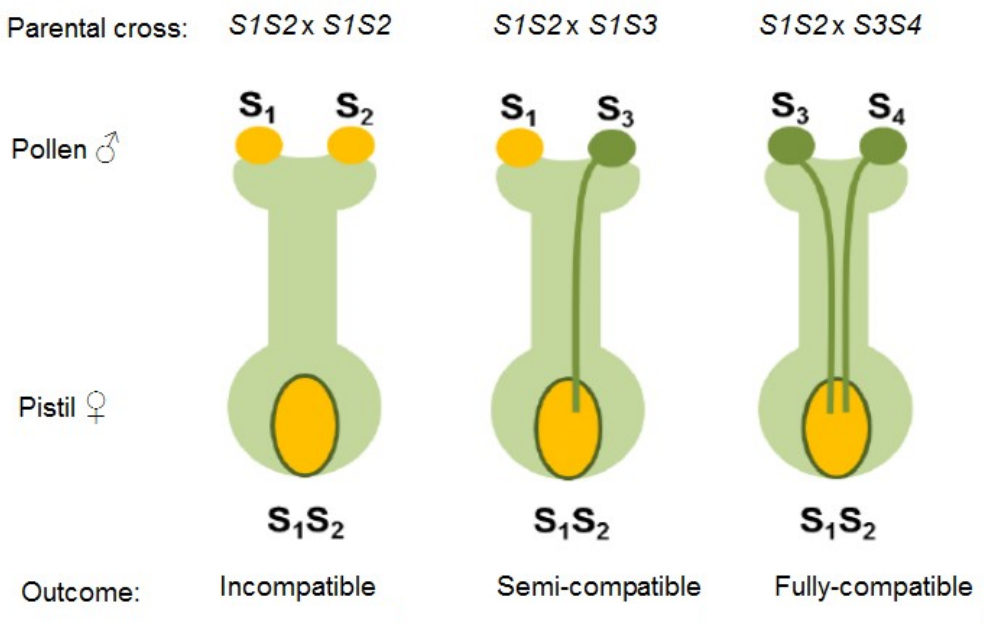


**Figure 2.2 Genetic control of the gametophytic self-incompatibility system showing incompatible, semi-compatible and fully-compatible crosses.** (Adapted and modified from McClure and Franklin-Tong (2006).

The advent on the *S*-RNAse was first identified in *Nicotiana alata* as a stylar basic glycoprotein for which the gene co-segregated with the pistil *S*-allele (Anderson *et al.*, 1986). The *S*-RNases are responsible for both pistil recognition and rejection activities of the *S*-haplotype (Wheeler *et al*. 2001; McClure *et al*., 1989, Lee *et al*., 1994). The *S*-RNases and analyses of structural genes have been explored extensively in cultivated fruit-bearing species, including apple (Ushijima *et al.*, 1998; Broothaerts, 2003). The sequence of the *S*-RNase gene (Figure 2.3) is composed of five conserved regions; C1, C2, C3, RC4 (*Rosaceae*-specific RC4) and C5, and a single intron located within the relative hyper-variable (RHV) region. The RHV region is located between the C2 and C3 regions, and it plays a critical role in determining allele-specific *S*-RNase activity (Long *et al.*, 2010; Minamikawa *et al.*, 2010)(Ushijima *et al*., 1998; Ishimizu *et al*., 1998; Long *et al*., 2010; Sassa *et al*., 2010).

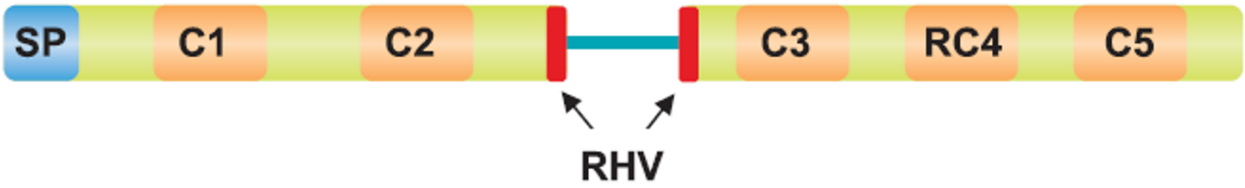


**Figure 2.3 Schematic representation of *S*-RNase sequence in *Malus*.** SP denotes signal peptide, C1, C2, C3, RC4 and C5 denote conserved regions. RHV is the Rosaceae hypervariable region of the S-RNase. The intron is located inside the RHV and represented by the blue horizontal bar. (Adapted and modified from Yamane and Tao, 2009).

The pollen *S*-determinant SFB/SLF genes were discovered almost 15 years after the pistil *S*-gene, *S*-RNase (Tao and Lezzoni, 2010). In *Rosaceae*, the *S*-locus was first sequenced in *Prunus species*, where a single F-box protein, SLF (*S*-locus F-box) or SFB (*S*-haplotype-specific F-box protein) were determined as the candidate genes for the pollen *S*-determinant (Entani *et al*., 2003; Ushijima *et al*., 2003; Zhang *et al*., 2007). Later studies in *Malus* and *Pyrus* identified multiple F-box proteins, called *S*-locus F-box brothers (SFBB), as good candidates since they exhibit *S*-haplotype-specific polymorphisms at the amino-terminal domain (Sassa *et al*., 2007; Minamikawa *et al*., 2010; Gu *et al*., 2015). For apple (*Malus domestica*), two SFBBs were identified around the *S*-RNases $S_3$ and $S_9$ genomic regions, and

referred to as MdSFBB3 and MdSFBB9, respectively (Sassa *et al*., 2007; 2010). Studies on pollen-derived cDNA of Japanese pear, *Pyrus pyrifolia* 'Kosui' ($S_4S_5$), described three (α; β; γ) SFBBs associated to each of the *S*-haplotypes $S_4$ and $S_5$: PpSFBB4-α, PpSFBB4-β, and PpSFBB4-γ, and PpSFBB5-α, PpSFBB5-β, and PpSFBB5-γ (Ushijima 2001; 2003; Okada, 2015). These SFBBs showed linkage to the *S*-RNase, *S*-haplotype-specific sequence divergence, and pollen-specific expression, and are considered good candidates for the pollen *S*-gene in apple and Japanese pear (Sassa *et al.*, 2007)(Sassa *et al*., 2007).

Kobel *et al*. (1939) as cited by Brown (1992), was the first to unravel the underlying *S*-genotype phenomena in apple by undertaking pollination studies which involved microscopic evaluation of pollen-tube growth, thereby distinguishing between fully compatible, semi-compatible and incompatible crosses. Based on these microscopic observations, the first eleven *S*-alleles in apple, $S_1$ to $S_{11}$ were determined and these findings resulted in resolving *S*-genotypes of 14 diploid and 12 triploid apple varieties (Brown, 1992). This was followed by Komori *et al*. (2000) who identified 10 *S*-alleles amongst Japanese apple cultivars, and assigned letter symbol $S_a$ to $S_i$ and $S_z$ and reported their correspondence to four of Kobel's *S*-alleles. Around that time, biochemical and molecular methods were developed and Sassa *et al*. (1996) discriminated S-alleles ($S_a$ to $S_f$) through characteristic migration patterns of gene products. Several studies followed that developed molecular diagnostic techniques for the identification of *S*-alleles based on allele-specific PCR amplification and restriction digestion (Broothaerts *et al.*, 1995; Janssens *et al.*, 1995). Since then, more *S*-alleles have been identified in *Malus* and renumbered (Broothaerts, 2003; 2004; Matsumoto *et al*., 2003; Kim *et al*., 2006) and many apple cultivars have been genotyped using *S*-allele specific primers (Broothaerts *et al*. 1995; Sassa *et al*. 1996; Verdoodt *et al*. 1998). This resulted in *S*-genotypes for many apple cultivars including; 'Golden Delicious' ($S_2S_3$), 'M9' ($S_1S_3$) 'McIntosh' ($S_{10}S_{25}$), 'Telamon' ($S_3S_{10}$), 'Trajan'($S_2S_{25}$), 'Tuscan' ($S_5S_{10}$), 'Northern Spy'($S_1S_3$) have been identified (Bošković and Tobutt, 1999; Broothaerts *et al*., 2004; Dreesen *et al*., 2010). To date, the *S*-locus has been mapped on linkage group 17 of the apple genome (Maliepaard *et al.*, 1998; Fernández-Fernández *et al.*, 2008).

### 2.3.3 Apple scion cultivars and rootstocks

More than 10,000 diverse apple cultivars exists worldwide, with only a few dominating the global economic apple market (Janick *et al.*, 1996; Hummer and Janick, 2009). In South Africa, the most commercially dominant apple cultivars (named in order of importance) are 'Golden Delicious', 'Granny Smith', 'Royal Gala', 'Topred', 'Starking', 'Cripp's Pink', 'Pink Lady', 'Rosy Glow', 'Fuji', 'Braeburn', 'Cripp's Red', 'Oregon Spur', 'Kanzi', 'Honeycrisp', and 'African Carmine'(DAFF, 2018; HORTGRO, 2018). Although apples can be propagated quite easily from seed, the resulting seedlings are extremely variable in vigour, growth habit and fruit characteristics due to the heterozygosity resulting from the self-incompatibility system (Jackson, 2003; Costes and García-Villanueva, 2007).

Traditionally apple rootstocks are propagated by seeds or vegetatively by layering, stooling or cuttings (Webster, 1995; Petri *et al.*, 2019). Apple seedlings are still used as rootstocks in some parts of the world (Janick *et al.*, 1996) but they typically make full sized trees, which are too vigorous for modern high-density apple orchards. Also, the performance of the scion cultivar grown on seedling rootstocks can be highly variable and unpredictable (Tworkoski and Miller, 2007; Petri *et al.*, 2019).

Clonal rootstocks have been used in the propagation and cultivation of the domesticated apple for more than a century (Pilcher *et al.*, 2008). The use of dwarfing rootstocks, which confer scion precocity, has reduced the generational intervals for apple trees to between three and five years. Clonal rootstocks are propagated vegetatively (asexually) through cuttings, stooling or layering methods. The clonal rootstocks are generally favoured as they are more uniform and may be chosen for certain desirable characters such as aphid resistance, cold hardiness, good soil anchorage, reduced tree vigour and reduced suckering (Fallahi *et al.*, 2002). Depending on the influence on the growth of the tree, rootstocks are classified as dwarf, semi-dwarf or vigorous. A dwarfing rootstock produces trees of about 15-30% the size of trees on apple seedlings (Webster, 2002; Costes and García-Villanueva, 2007). Dwarfing rootstocks have become a key factor in improving the efficiency of commercial apple production. They can be planted at a much higher densities than full size trees and can greatly improve the yield per acre of land, especially in the early years of the orchard. Moreover, the

resultant dwarf trees are advantageous in that they are easier to prune, pick and spray (Wang *et al.*, 2019).

A large number of apple rootstock cultivars are available. The most widely used range of apple rootstocks is the Malling (M) series classified at the East Malling Research Station in the United Kingdom and the Malling-Merton (M.M) series derived from a combined breeding effort for resistance to woolly aphids. The rootstocks in the Malling series are named and assigned numbers from 'M.1' to 'M.27' while the MM series ranges from 'M.M 101' to 'M.M 116', though some selections were not released (Jackson, 2003). 'Malling 9' is the most widely used rootstock worldwide. It was originally called 'Jaune de Metz' and was discovered in France as a single plant in 1879 (Fallahi *et al.*, 2002). 'Malling 9' is a clonal dwarf rootstock, much in demand for producing small trees for commercial high density plantations. Its characteristics are an early cropping and amelioration of fruit quality but it is sensitive to woolly apple aphid (*Eriosoma lanigerum*) (Webster, 2002). However, 'M.9' has not been utilised that often in South Africa due to hot soils, although some trials exist, and it has also been used as an interstem.

## 2.4 Hybrid incompatibility and plant immunity

During evolution, ancestral species may diverge into several species that become genetically isolated from one another and develop a reduced ability for hybridization; referred to as hybrid incompatibility (Orr, 1996; Rieseberg and Willis, 2007; Rieseberg and Blackman, 2010). Hybrid incompatibility in the form of hybrid necrosis arises from the deleterious interaction or negative epistatic interactions between alleles that arose from independent genetic backgrounds as explained by the Dobzhansky-Muller model (Orr, 1996; Orr and Presgraves, 2000; Bomblies et al., 2007). Auto-immunity or hybrid necrosis is synonymous with a seedling phenotype typically characterised by tissue necrosis, wilting, yellowing, chlorosis, dwarfism, reduced growth rate, and in some cases lethality (Bomblies and Weigel, 2007). Along with hybrid necrosis, hybrid sterility, and hybrid inviability are other postzygotic reproductive barriers that arise from epistatic deleterious interactions between alleles from two parents (Orr and Presgraves, 2000; Bomblies and Weigel, 2007; Gross, 2007; Chen et al., 2016). Also, genetic incompatibilities, resulting in segregation distortion is

considered fundamental to postzygotic isolation leading to speciation (Bomblies *et al.*, 2007; McDermott and Noor, 2010).

Plant defense mechanisms detect pathogen-associated molecular patterns (MAMP's) or pathogen- microbe-associated molecular patterns (PAMP's) or damage-associated molecular patterns (DAMP's) through the pattern recognition receptors (PRRs) (Thomma *et al.*, 2011; Macho and Zipfel, 2014; Hou *et al.*, 2019). The PRRs are usually plasma membrane-localised receptor like kinases (RLKs) or receptor-like proteins (RLP) with extracellular functional domains allowing MAMP/DAMP perception (Macho and Zipfel, 2014; Hou *et al.*, 2019). The defense response in many host species can be triggered by a broad range of structurally diverse PAMPs (elicitors), including direct detection of molecules of microbial origin such as bacterial flagellin and elongation factor Tu, fungal chitin, and lipopolysaccharides (Segonzac and Zipfel, 2011; Macho and Zipfel, 2014; Park *et al.*, 2014).

Signaling cascade is a highly evolved intricate network of signaling protection entities that a plant employs in detecting and fighting biotic and abiotic stresses (Jones and Dangl, 2006; Mhamdi and Van Breusegem, 2018). Reactive oxygen species (ROS) play important roles in many signal transduction pathways mediating responses to pathogen infection, abiotic stress, developmental regulation, and programmed cell death in different cell types (Karuppanapandian *et al.*, 2011; Schippers *et al.*, 2012; Schmidt and Schippers, 2015). Reactive oxygen species (ROS) include the superoxide radical anion ($O_2^{\bullet-}$), hydroperoxyl radical ($HO_2$), hydroxyl radical ($HO^{\bullet}$), hydrogen peroxide ($H_2O_2$) and singlet oxygen ($^1O_2$) (Mittler *et al.*, 2004, 2011; Karuppanapandian *et al.*, 2011).

Plants generate ROS by activating several enzymatic systems that have been recognized as cellular sources of the oxidative burst in the apoplast of plant cells. These include NADPH oxidases, cell wall peroxidases (class III secretory plant peroxidases), amine and polyamine oxidases, oxalate oxidases and quinone reductases (Karuppanapandian *et al.*, 2011; Sewelam *et al.*, 2016). Peroxidases have been proposed as alternative producers of ROS (Passardi *et al.*, 2005; O'Brien *et al.*, 2012). Peroxidases (class III peroxidase) catalyze the oxidoreduction of various substrates using $H_2O_2$.and are implicated in detoxification of ROS

and diverse biological functions, such as signal transduction, lignin biosynthesis, seed germination, senescence, and host-pathogen interactions (Passardi et al., 2005).

### 2.4.1 Hybrid incompatibilities in apple (*Malus*)

Hybrid incompatibilities have been previously reported in several progenies of apple (*Malus* spp), for which F1 hybrids from parents of normal habits exhibited morphological abnormalities, which are associated with deleterious traits. In 1939, Crane and Lawrence (as cited by Fernández-Fernández *et al.*, 2014) reported a cross of 'M.8' x 'M.9' and their reciprocals which exhibited F1 hybrids which were yellow and some albino in colour. All the albino seedlings died soon after germination. In contrast, some of the yellow seedlings died whilst the remaining ones persisted. At bud burst the leaves were reported intensely yellow, but changed to pale-green as the season advanced (Fernández-Fernández *et al.*, 2014).

In another cross of 'M.16' x 'M.9', Way *et al*. (1976) demonstrated seedlings which were characterised by whitish-yellow leaves and of low growth vigour. These seedlings were assigned the name 'virescence' (Fernández-Fernández *et al.*, 2014). In the same year, Alston (1976), reported seedlings of pale-green lethal, which were characterised by seedlings deficient in chlorophyll and died few weeks after germination.

Alston (1976), also reported three rare dwarf types (early, crinkle and sturdy dwarfs). Early dwarfs appeared four weeks after germination in small progenies and grew to no more than 120 mm in the first growing season. The plants had very short internodes and rarely survived harsh winter conditions. The early dwarfs were suggested to be controlled by two independent recessive genes. The sec, crinkle dwarf phenotype were characterised by small rounded crinkled leaves with normal internodes. The seedlings could grow between 300 to 600 mm over two years, and is attributed to control by recessive genes. As explained later, this phenotype may be the same as one of those studied in the current thesis. The third type, sturdy dwarf phenotypes had high numbers of branches and very short internodes (Alston, 1976). They were reported to have a long juvenile period. Sturdy dwarfs could grow up to one meter in height over three years. However, they are not as easily distinguishable from early dwarfs, and could only be pre-selected eight weeks after germination just before

planting in the field. Sturdy dwarfs were attributed to control by several recessive genes at least at two loci (Alston, 1976).

Thereafter, a study of Haniuda *et al*. (1985) (as cited by Fernández-Fernández *et al.*, 2014), reported a cross and backcross of 'M.9' x 'M.27'; 'M.13' x 'M.9', which exhibited F1 seedlings with whitish-yellow leaves in spring which were named 'pseudo-albino'. As the season changed to summer, the whitish-yellow leaves progressively changed to green-leaves. The pseudo-albino were suggested to be controlled by two loci.

Gao and van de Weg (2006), studied apple progenies, which exhibited F1 seedlings of poor vigour, most of which died three months after germination, while others persisted for a longer time. At germination, these seedlings had green cotyledons, epicotyls and leaves. As the growing season progressed, the seedlings showed retarded growth with extremely short internodes and leaves which were smaller than the cotyledons. These progenies exhibit lethal and sub-lethal genes at different developmental stages, for which *sl1* and *sl2* control lethality after and before germination, respectively, and *sl3* being sub-lethal. These seedlings abnormalities were similar to that of the early dwarf described by Alston (1976). Fernández-Fernández *et al.* (2014) studied a cross of 'M.27' x 'M.116', with F1 seedlings which were characterised with chlorotic leaves at germination or at bud break but changes to green-coloured leaves as the season progresses. These seedlings were also classified under 'virescent', similar to that reported by Way et al. (1976).

### 2.4.2 Hybrid incompatibilities in pear (*Pyrus*)

Recently, two types (type1 and type2) of hybrid necrosis which are characterised by stunted growth and lethality were observed at different developmental stages in the interspecific pear population of a cross between PEAR3 (*Pyrus bretschneideri* x *Pyrus communis*) and 'Moonglow' (*Pyrus communis*) (Montanari *et al.*, 2016). The type1 was characterised by leaves which were chlorotic and necrotic lesions, and died within one month after germination. The seedlings that survived had small leaves and had stunted growth which were less than 50 mm in height. On the contrast, the type2 hybrid necrosis developed normal at germination, and progressively began to cup downward. The leaves became chlorotic with

necrotic lesions, and plant growth development ceased within three months after germination (Montanari *et al.*, 2016).

## 2.5 Growth promoting hormones and dwarf mutants

Dwarf mutants have been identified in many species and have been investigated for their mode of inheritance and response to plant hormones (Ashikari *et al.*, 1999). Plant growth-promoting hormones such as gibberellins (GA) and brassinosteroids (BR) play an important role in plant height determination and organ expansion including root and stem elongation, rosette expansion, floral induction and anther development (Zhang *et al.*, 2008; Hedden and Thomas, 2012). There are many types of GA not have been identified, but only the bioactive forms are reported to be directly involved in growth regulation (Daviere and Achard, 2013). Molecular genetic studies of dwarf mutants in *Arabidopsis* (*Arabidopsis thaliana*) and other crop species revealed that the biosynthesis and signaling of GA and BR are the most important factors in determining plant height (Fujioka and Yokota, 2003; Yamaguchi, 2008). Numerous dwarf mutants have been reported in *Arabidopsis* such as *gai* (gibberellic acid insensitive), *rga* (repressor of ga1-3) and *shi* (short internodes) (Strader *et al*., 2004). The *GAI* and *RGA* are homologous genes that encode putative transcriptional regulators that repress GA signaling in *Arabidopsis* (Dill *et al.*, 2001). The semi-dwarf *gai* mutant encodes a mutant protein with a 17-amino acid deletion within the *DELLA* domain of the *GAI* protein which results in a semi-dwarf phenotype (Fu *et al*., 2001; Pent *et al*., 1999). The semi-dominant *gai* mutant was further characterised by germination failure, late flowering and male sterility, due to the defects in the *GAI* genes (Peng *et al.*, 1999). Therefore, *gai* is a negative GA-response regulator in *Arabidopsis*. The recessive *rga* mutant has been reported to suppress the phenotypic defects of the *Arabidopsis* GA biosynthesis mutant *ga1-3* (Dill *et al*., 2001). Silverstone *et al*. (1999) also reported that the defects in stem elongation, flowering time and leaf trichome initiation are suppressed by *rga* mutant which is indicative that *RGA* is a negative regulator of the GA signal transduction pathway. Another *Arabidopsis* protein, *SHI*, was reported to be a negative regulator of GA-induced cell elongation (Fridborg *et al.*, 2001). In the semi-dominant *shi* mutant, a transposon insertion confers over-expression of the *SHI* gene, resulting in the semi-dwarf phenotype. The *shi* mutant was characterised by

short hypocotyl, very short internodes, darker-green narrowed leaves, reduced apical dominance and late flowering (Fridborg, 2000). According to Fridborg *et al*. (2001), application of high doses of exogenous GA did not correct the dwarf phenotype, which is evident that *shi* was affected negatively in GA responsiveness.

## 2.6 Apple breeding and improvement

The first controlled pollination apple breeding was initiated early in the nineteenth century by Thomas Andrew Knight (1759–1835) in the UK, who crossed different apple cultivars and selected superior phenotypes with known parentage (Janick *et al.*, 1996; Pereira-Lorenzo *et al.*, 2009). The success of apple breeding improvement largely depends on the available genetic diversity. To date, controlled pollination is still the main approach followed by apple breeders. It is the most effective way of increasing the frequency of the desirable alleles due to the relatively high additive variance in most of the traits (Janick *et al.*, 1996)

In South Africa, the biggest apple breeding programme is conducted by the Agricultural Research Council (ARC) in the Cultivar Development (Breeding and Evaluation) Division of ARC Infruitec-Nietvoorbij. Its aim is to develop a range of improved cultivars which are well adapted to South African climates, are disease resistant, precocious and of good productivity with attractive appearance and good fruit quality. There is also an interest in developing rootstock ranges which are resistant to pest and diseases and have acceptable horticultural traits, but currently there is no action on rootstock breeding at the ARC.

In general, apple is not an easy crop for breeding due to the long juvenile phase of about 3 to 10 years, large field space requirement for planting and growing of seedlings (Gardiner *et al.*, 2007). Maintenance of these plantings also limits large-scale establishment and evaluation of apple breeding materials (Maliepaard *et al.*, 1998). Breeding of new apple cultivars is a very long and tedious process as it usually takes at least 20 years from cross pollination, seedling establishments up to the release of an apple cultivar (Gessler and Patocchi, 2007). Conventional plant breeding has been used for decades in developing new apple varieties using deliberate interbreeding of closely or distant related species (Janick *et al.*, 1996). However, conventional plant breeding can no longer sustain the global demand, and over the

years apple breeding was enhanced through amongst other things molecular marker technologies (Sansavini *et al.*, 2004). The latter approach allows early selection of plants displaying desired characteristics at a seedling stage. Ongoing improvement of marker detection systems and the identification of molecular markers linked to traits of interest can aid the introgression of more favourable genes into new cultivars while improving on selection strategies.

## 2.7 Marker-assisted selection (MAS)

Marker-assisted selection (MAS) has gained popularity as a powerful tool to efficiently select plants carrying genomic regions of traits of interest based on molecular markers (Collard *et al.*, 2005). This approach is used as an indirect selection process based on the molecular marker-trait association rather than the actual gene-selection. In practice, selection of desirable genotypes can be made at the early stages of plant development (e.g. seedling stage) thus replacing phenotype with genotype-based selection (Collard and Mackill, 2008; Jannink *et al.*, 2010). This is particularly valuable in apple breeding as traits related to production generally only becomes apparent after the plant has overcome the hampering juvenile phase, which can last for six to ten years (Janick *et al.*, 1996). In apple, MAS has mostly been focused on disease resistance with several resistance apple scab, *Venturia inaequalis* genes (*Vf*) and woolly apple aphid, *Eriosoma lanigerum*, *Er1* and *Er3* been incorporated into breeding lines (Bus *et al.*, 2000, 2008; Kumar *et al.*, 2011). Additionally, MAS of other traits such as flavour, firmness, columnar growth habit, and skin colour has been performed (Takos *et al.*, 2006; Zhu and Barritt, 2008; Petersen and Krost, 2013).

## 2.8 Molecular Genetics in Apple

Modern fruit breeding is dependent on molecular markers for the rapid and accurate assessment of germplasm, trait mapping and marker-assisted selection. Molecular markers are identifiable allelic variations that distinguish individuals, populations and species in terms of nucleotide differences. These variations may have direct effects on the phenotype but are more often linked to a trait of interest (Nadeem *et al.*, 2018). Typically, molecular markers do not represent target genes themselves, but are rather tags or landmarks in the genome used to

keep track of a chromosome region (Mahajan and Gupta, 2012). During the 1980s and 1990s, various types of molecular markers were developed. However, these molecular markers have many disadvantages (see Table 2.1 for comparison) and currently microsatellites and high throughput SNPs predominate applications in modern plant genetic analysis.

**Table 2.1 Comparison of most commonly used molecular marker systems**

| Characteristics | RFLPs | RAPDs | AFLPs | SSRs | SNPs |
|---|---|---|---|---|---|
| **DNA required (µg)** | 10 | 0.02 | 0.5 - 1.0 | 0.05 | 0.05 |
| **DNA quality** | high | high | moderate | moderate | high |
| **Polymorphic loci** | 1 - 3 | 1 - 50 | 20 - 100 | 1 - 3 | 1 |
| **PCR based** | no | yes | yes | yes | yes |
| **Ease of use** | difficult | easy | easy | easy | easy |
| **Amenability** | low | moderate | moderate | high | high |
| **Reproducibility** | high | unreliable | high | high | high |
| **Development cost** | low | low | moderate | high | high |
| **Cost per analysis** | high | low | moderate | low | low |

### 2.8.1 Microsatellites or Single Sequence Repeats (SSRs)

Microsatellites, commonly known as single sequence repeats (SSRs), are the smallest class of simple repetitive DNA sequences. They consists of two to six nucleotides that are tandemly repeated e.g. mono-, di-, tri- and tetra-nucleotide repeat sequences (Schlötterer, 2000). Microsatellites are highly polymorphic and ubiquitously distributed throughout eukaryotic and prokaryotic genomes (Senan *et al.*, 2014). Microsatellites can be found in both coding (genic-SSRs or EST-SSRs) and non-coding regions (genomic-SSRs), but are most frequent in non-coding regions due to negative selection against frameshift mutations in coding regions (Gianfranceschi *et al.*, 1998; Varshney *et al.*, 2005). The nucleotide repeat units are flanked by specific DNA sequence unique for a given locus from which locus-specific primers can be designed. These primers can be used to amplify the target sequences by PCR and reveal specific alleles differing by variable base pair sizes (Tautz, 1989; Moniruzzaman *et al.*, 2016). SSRs are abundant, locus-specific and amenable to various high-throughput genotyping methods. The co-dominant nature of SSRs allows for the distinction between homozygous and heterozygous genotypes (Miah *et al.*, 2013; Vieira *et al.*, 2016).

Despite microsatellite's usefulness, many aspects of microsatellite evolution are still not clear, particularly the implications of microsatellite mutation and diversification on genome dynamics (Teneva *et al.*, 2014). Microsatellite loci are inherently unstable with high level of polymorphism (e.g.$10^{-2}$ to $10^{-6}$ mutations per locus per generation) attributable to mutation mechanisms; mainly strand-slippage mispairing during DNA replication or repair (Levinson and Gutman, 1987; Viguera *et al.*, 2001; Oliveira *et al.*, 2006), and minor contributions from mismatched recombination (gene conversion and unequal crossing-over (Schlötterer, 2000; Leclercq *et al.*, 2010).

Genotyping scoring errors are a major drawback of microsatellites and may arise due to: i). Null alleles, occurs mostly due to primer annealing site mutation (point mutation), resulting in failure of amplification during the PCR reaction or due to poor cross-species amplification. Rico *et al*. (2017) stated that null alleles can also occur via segmental aneuploidy, where one chromosome has a deletion containing the primer binding site. Therefore, scoring of null alleles may lead to biased estimates of the allelic frequencies, especially underestimation of heterozygosity (González-Robles *et al.*, 2016; Rico *et al.*, 2017). ii). Homoplasy, 'size homoplasy' refers to electromorphs which are identical in state (same size), but may not be identical by descent (Estoup A *et al.*, 2002; Germain-Aubrey *et al.*, 2016). The size similarity may arise due to convergence, reversion or parallism. iii). Stutter bands, occur as artifacts generated during PCR amplification and on account of them, actual allele sizes are difficult to distinguish from stutter bands (Rédei GP, 2008; Hosseinzadeh-Colagar *et al.*, 2016). The selection of the correct band or peak, size determination of the fragments, and interpretation of the band profiles may mislead interpretation, thus resulting in false positives (Ellegren, 2004).

Despite these shortfalls microsatellites have a broad range of applications, and another great advantage of SSR markers is the possibility of multiplexing, which allows several SSRs to be combined in one PCR reaction. This significantly decreases the costs of sequencing analysis (Li *et al.*, 2018). Microsatellites have proven very informative for diversity study, cultivar identification, and in the characterisation of qualitative and quantitative traits. To date, more than 1,200 *Malus* SSR markers have been developed and mostly published (Liebhard *et al.*, 2002; Silfverberg-Dilworth *et al.*, 2006). Most of these SSRs are represented on the Genome

Database of Rosaceae (www.rosaceae.org) and HiDRAS SSR Database (http://www.hidras.unimi.it/HiDRAS-SSRdb) (Jung *et al.*, 2008).

### 2.8.2 Single Nucleotide Polymorphisms (SNPs)

In recent years, SNPs have gained popularity in a number of breeding industries. Single nucleotide polymorphism are the latest generation of DNA markers, and constitute the most frequent type of genetic variation (Chagné *et al.*, 2007; Rasal *et al.*, 2017; Rasheed *et al.*, 2017). Single nucleotide polymorphisms are DNA sequence variations occurring where two DNA sequences from homologous chromosome differs by a single nucleotide base (Gupta *et al.*, 2001) and are largely mutations resulting from DNA replication errors (Edwards *et al.*, 2007). They are the most abundant genetic variations and distributed evenly in high frequencies throughout the genome of most animal and plant species (Ganal *et al.*, 2009; Deschamps *et al.*, 2012). For a variation to be considered a SNP, it must occur at a frequency of a least 1% or greater in a population, since a frequency of less the 1% is generally considered a point mutation (Brookes, 1999; Rafalski, 2002).

SNPs are classified based on the types of mutational events and may arise in three different mutational forms (Huq *et al.*, 2016); i) transitions, in which the nucleotide is replaced by a nucleotide of the same class, purine to purine or pyrimidine to pyrimidine (eg. C→T or A→G), ii) transversions are defined by a change in nucleotide from a different different class, purine to pyrimidine and pyrimidine to purine (eg. C→G, A→T, C→A, or T→G) and/or, iii) small INsertions/DELetions (INDELs) (Appleby *et al.*, 2009). SNPs may additionally fall within non-coding and coding regions. Non-coding SNPs occur in regions of DNA which does not code for proteins such as intergenic and intron regions, whereas, coding-SNPs occur in exons, and both have the potential of altering protein function through the creation of stop codons (Edwards *et al.*, 2007; Rasheed *et al.*, 2017).

Of all the molecular marker types used to date, SNPs are the most abundant, robust and feasible for high-throughput genotyping, especially through SNP arrays which allow for simultaneous screening of hundreds and thousands of loci and automated allele calling (Ganal *et al.*, 2012; Troggio *et al.*, 2013; Vanderzande *et al.*, 2017). However, SNPs are bi-allelic in

nature, which could make them less informative when compared with multi-allelic microsatellite markers, but their abundance and uniform genome distribution overcomes this limitation (Jehan and Lakhanpaul, 2006; Edwards *et al.*, 2007). SNPs are also considered to be evolutionarily stable with low mutation rate (Gupta *et al.*, 2001; Edwards *et al.*, 2007; Rasheed *et al.*, 2017).

Recent advances in technologies for high-throughput SNP based genotyping have since improved, and large numbers of SNPs are being identified at reasonable cost for different purposes including cultivar identification, construction of high-resolution genetic maps, fine mapping of quantitative trait loci and genome-wide association studies from a wide range of population (Rafalski, 2002; Houston *et al.*, 2014). The advent of sequencing the genome arose in 2000 with the sequencing of the human genome, which was found to contain about 1.42 million SNPs at an average of one SNP per 1.9 Kbp (Sachidanandam *et al.*, 2001). Studies and applications of sequence diversity have since been performed for a range of plant species and these have indicated that SNPs appear to be abundant in plant systems with one SNP every 100–300 bp (Rafalski, 2002; Appleby *et al.*, 2009). The number of SNPs present in the genome varies among species.

SNPs were first applied in apple in 2008, with the development of a set of SNP markers proposed by Chagne *et al*. (2008) who highlighted the improved genomic coverage and ease of genotyping. The wide application of SNP markers was made possible by the availability of the apple 'Golden Delicious' reference genome (Velasco *et al.*, 2010), which became a milestone for the apple breeding industry, and led to the identification of more than 2 million SNPs (2,113,120) at an average of four SNPs per 1,000 bp (Chagné *et al.*, 2012). More genotyping arrays, including the apple IRSC 8K, Illumina Infinium 20K and Axiom apple 480K SNP arrays, were anchored to the apple reference genome, permitting the identification of genomic regions underlying genes of interests and potential candidate genes (Chagné *et al.*, 2012; Bianco *et al.*, 2014, 2016).

In 2012, the International RosBREED SNP Consortium (IRSC) array, a 8K SNP array, was developed based on genic SNPs detected in 27 apple cultivars and consequently used to construct a dense linkage map of almost 1000 markers for 'Royal Gala' x 'Granny Smith'

(Chagné *et al.*, 2012). This array contains 7,867 *Malus* SNP markers and 921 SNPs derived from *Pyrus* (Chagné *et al.*, 2012; Montanari *et al.*, 2013) and has already been applied widely. This includes screening bi-parental populations which has greatly facilitated the development of high density linkage maps in apple (Antanaviciute *et al.*, 2012; Troggio *et al.*, 2013; Clark *et al.*, 2014), clarification of parentages (Pikunova *et al.*, 2014; Howard *et al.*, 2017), and identifying genomic regions differing between cider and dessert apples (Leforestier *et al.*, 2015). Following this, a higher density Illumina Infinium® 20K SNPs array, containing about 3.7K SNP markers from the IRSC 8K array and newly developed SNP markers from 14 apple genotypes and two double-haploids from 'Golden Delicious' (Bianco *et al.*, 2014) was developed. This array allowed for the detection of QTLs for several apple quality traits (Falginella *et al.*, 2015; Sun *et al.*, 2015; Di Guardo *et al.*, 2017; Farneti *et al.*, 2017). The largest and most highly saturated apple SNP array is the Axiom 480K, developed within the EU-FruitBreedomics project, built from the re-sequencing of 63 different apple cultivars (Bianco *et al.*, 2016). This array contains 487,249 SNPs evenly distributed over the 17 apple chromosomes and is the most promising for genome-wide association (GWA) studies in apple. It enlarges the sequence database and furthers the depth of apple genotype descriptions (Bianco *et al.*, 2016), and has already allowed for a number of applications including GWA of the genetic control of flowering and ripening periods as well as evaluating natural diversity and the presence of non-functional naturally occurring alleles of MdMLO19 in apple germplasm (Urrestarazu *et al.*, 2017); Pessina *et al.*, 2017). Single nucleotide polymorphism array platforms have also been developed in other *Rosaceae* species including strawberry (Bassil *et al.*, 2015), cherry (Peace *et al.*, 2012), peach (Verde *et al.*, 2012), and pear (Li *et al.*, 2019). SNPs have proven to be an important resource facilitating the production of dense genetic maps and QTL mapping of important traits, genome-wide association analysis, pedigree- based analysis, and genomic selection (Bianco *et al.*, 2014, 2016; Urrestarazu *et al.*, 2017).

### *Genetic Linkage Mapping*

The advent of molecular markers has greatly facilitated the construction of genetic linkage maps, and detecting positions of genes on chromosomes and respective assignment to linkage groups (LGs) (Collard *et al.*, 2005; Semagn *et al.*, 2006). Genetic linkage maps play a major role in clarifying the genetic control of important traits and serve as a prerequisite for development of DNA-based diagnostic tools for marker assisted breeding (Di Pierro *et al.*, 2016).

Linkage mapping is mainly based on the genetic principles of segregation and recombination during meiosis in accord with Mendel's law of independent assortment (Semagn *et al.*, 2006). During meiosis, homologous chromosomes crossover and exchange DNA segments, forming hybrid chromosomes, a process called recombination. In a case where two loci are located on different chromosomes or far apart on the same chromosome, their transmission of alleles will be random and therefore assort independent of each other, and will be unlinked. On the other hand, if two loci are situated in close proximity on the same chromosome their transmission of alleles tend to be co-inherited, exhibiting linkage in either coupling or repulsion phase (Semagn *et al.*, 2006; Portin, 2014).

Linkage between paired molecular markers is calculated by the use of the logarithm of the Odds (LOD) score, a measurement of likelihood of linkage which used as a threshold of linkage significance (Stam, 1993). In general, LOD scores equals to the log10 of the probability that two loci are linked over the probability that two loci are unlinked (Risch, 1992; Stam, 1993; Morton, 1996; Nyholt, 2000; Xu, 2013). Evidence for linkage is generally indicated by a LOD score of three and above as the significance threshold. This affirms that the likelihood of the molecular marker linkage is at least 1000 times more probable (1000:1) favouring linkage (Collard *et al.*, 2005; Xu, 2013; Chandra and Pandey, 2017). The estimation of recombination frequencies may be complex when large numbers of individuals and molecular markers are investigated, and are therefore converted to mapping distances, as the two variables are not linear. To handle such calculations, various computer programs exists, one being JoinMap (Jansen *et al.*, 2001; Van Ooijen and Voorrips, 2006; Van Ooijen, 2011). The two most commonly used mapping functions in JoinMap are Haldane (1919) and

Kosambi (Kosambi, 1944), and are expressed in centiMorgans (cM), in which 1 unit of crossing-over corresponds to a 1% frequency of recombination events in 100 gametes (1cM=1% recombination) (Stam, 1993)(Stam, 1993; Liu, 1998; Kumar, 1999). The Haldane mapping function assumes there is no interference which would increase or decrease the equally probable crossover events between loci, and is thus not accurate for longer marker distances. The Kosambi mapping function incorporates all the interferences in genetic distance estimation, making it a more sensitive algorithm for linkage map construction (Collard *et al*., 2005; Danzmann and Charbi, 2007; Huehn, 2010). Marker ordering analyses are then employed to obtain the best molecular marker order relevant to the mapping population. The older JoinMap versions uses a regression mapping algorithm (Stam 1993), which is based on the sequential addition of molecular markers in a pairwise and goodness-of-fit manner. However, it cannot effectively analyse data sets with over 50 molecular marker loci mapped to a single linkage group (Stam, 1993; Van Ooijen and Voorrips, 2001). The newer JoinMap version has developed a commonly used multipoint maximum-likelihood algorithm which incorporates the complexities of mapping extremely large data sets (Jansen *et al*., 2001; Van Ooijen, 2011). The schematic genetic linkage maps are drawn using MapChart software (Voorrips, 2002).

In apple, factors such as self-incompatibility, high levels of heterozygosity and the long juvenile phase (Peace and Norelli, 2009) pose limitations in the generation of segregating populations such as filial generation 2 (F2), backcrossing (BC) or recombinant inbred lines (RIL), which are commonly used for the mapping in other plant species. Therefore, linkage mapping in apple has traditionally been conducted in filial generation 1 (F1) using the two-way pseudo-testcross (Grattapaglia and Sederoff, 1994; Maliepaard *et al*., 1998; Gessler and Patocchi, 2007). The pseudo-testcross method overcomes map construction barriers caused by parental heterozygosity. In an F1 population, up to four alleles segregate at one locus, resulting in a possibility of nine genotypic combinations: <abxcd>, <abxac>, <abxab>, <abxaa>, <abxcc>, <aaxab>, <ccxab>, <aaxbb>, <aaxaa>. Of these nine genotypes, five marker types can be defined for the construction of a genetic linkage map (segregation of co-dominant marker types in F1 is summarised in Table 2.2 (Van Ooijen and Voorrips, 2017).

**Table 2.2. Segregation of co-dominant markers in outbred F1 crosses**

| Parental genotypes | Expected genotypes | Expected segregation | Mapping usefulness |
|---|---|---|---|
| ab x cd | ac, ad, bc, bd | 1:1:1:1 | Anchor marker (segregating and mappable on both parents) |
| ef x eg | ee, eg, ef, fg | 1:1:1:1 | Anchor marker (segregating and mappable on both parents) |
| hk x hk | hh, hk, kk | 1:2:1 | Marker phase must be known to be mappable |
| lm x ll | ll, lm | 1:1 | Segregating in first parent only |
| nn x np | nn, np | 1:1 | Segregating in sec parent only |

A number of apple genetic linkage maps have been developed and published. More than 1200 molecular markers are distributed across the 17 linkage groups (Liebhard *et al.*, 2002; Silfverberg-Dilworth *et al.*, 2006; Celton *et al.*, 2009). The first apple linkage map was constructed by Hemmat *et al.* (1994). This linkage map consisted mainly of isoenzyme, RFLP and RAPD markers. The sec linkage map was developed by Conner *et al.* (1997) and also included isoenzymes and RAPDs. Thereafter, the first detailed genetic map comprising of 17 linkage groups was constructed by Maliepaard *et al.* (1998) using isozyme, RAPD, RFLP, SSRs and SCAR (sequence characterised amplified regions) markers. This map enabled the positioning of the scab resistance genes, *Venturia inaequalis* (*Vf*), rosy leaf curling aphid (*Sd1*), fruit acidity gene (*Ma*) and self-incompatibility locus, *S*. A more dense apple genetic linkage map was then constructed by Liebhard *et al.* (2002, 2003) to which Silfverberg-Dilworth *et al.* (2006) added 156 SSR markers. Thereafter, several high density linkage maps for apple were constructed using informative SSR markers (Naik *et al.*, 2006; Fernández-Fernández *et al.*, 2008; Celton *et al.,* 2009; Pattochi *et al.*, 2009). Moreover, an integrated map was developed from six apple populations and used for the anchoring of the whole apple genome sequence (Velasco *et al.*, 2010) leading to additional linkage maps summarised in Table 2.3. (Antanaviciute *et al.*, 2012; Fernández-Fernández *et al.*, 2012, 2014; Clark *et* Gaur *et al.*, 2015; Sun *et al.*, 2015).

**Table 2.3 Genetic linkage maps of apple (*Malus*)**

| Cross | Population size | Marker types | Trait | References |
|---|---|---|---|---|
| Rome Beauty x White Angel | 56 | Isoenzyme, RFLP, RAPD | Pl-w | Hemmat *et al*., 1994 |
| Wijcik McIntosh x NY 75441-67 | 114 | Isoenzyme, RAPD | Rf, Vf, Co, Ma | Conner *et al*., 1997 |
| Wijcik McIntosh x NY 75441-58 | 172 | Isoenzyme, RAPD | | Conner *et al*., 1997 |
| Prima x Fiesta | 152 | Isoenzyme, RFLP, RAPD, AFLP, SCAR | Vf, Sd-1, Ma, S | Maliepaard *et al*., 1998 |
| Fiesta x Discovery | 112 | RAPD, SSR | | Liebhard *et al* ., 2002 |
| Fiesta x Discovery | 267 | RAPD, AFLP, SSR, SCAR | | Liebhaard *et al* ., 2003 |
| Discovery x TNR10-8 | 149 | Isoenzyme, AFLP, SSR | Vg, scab QTL, RGH | Calenge *et al*., 2004; 2005 |
| Telamon x Braeburn | 257 | AFLP, SSR | | Kenis and Keulemans, 2005 |
| Fiesta x Totem | 85 | SSR, SCAR | Vf, Pl-2, Co, Rt, Gfc, ETR1 | Fernandez-Fernandez *et al*., 2008 |
| M.9 x R.5 | 94 | RAPD, SSR, SCAR, SNP | | Celton *et al*., 2009 |
| Co-op 17 x Co-op 16 | 142 | EST-SSR | | Han *et al*., 2011 |
| M.27 x M.116 | 140 | SSR | S | Fernandez-Fernandez *et al*., 2012 |
| JM7 x Sanashi 63 | | EST-SSR, AFLP | S | Moriya *et al*., 2012 |
| M.27 x M.116 | 188 | SSR, SNP | Vir | Fernandez-Fernandez *et al*., 2014 |

### 2.8.3 High-throughput molecular approaches in apple

Currently, three genomes in apple (*Malus domestica*, 'Golden Delicious') have been sequenced and released publically. In 2010, the reference genome sequence of apple 'Golden Delicious' was made publicly available for the first time, and greatly accelerated apple genomics research (Velasco *et al.*, 2010). Roughly 13 billion nucleotides of diploid 'Golden Delicious' apple was sequenced and assembled using Sanger and Roche 454 technologies. Though the sequenced nucleotides provided a 17-fold genome coverage, only 742.3 million nucleotides were assembled while roughly 90.2% of the genes could be assigned to chromosomes (Velasco *et al.*, 2010). Based on this reference genome, it was also revealed that large chromosomal regions are duplicated (Troggio *et al.*, 2012). The putative gene content found in apple was 57,386 and reported the highest amongst the fruit crop species studied at the time.

In 2016, an improved hybrid *de novo* genomic assembly of *Malus domestica* 'Golden Delicious', was obtained from 76 Gb (~102 × genome coverage) Illumina HiSeq data and 21.7 Gb (~29 × genome coverage) from PacBio data (Li *et al.*, 2016). The final draft genome is approximately 632.4 Mb, representing about 90% of the estimated genome. The contig N50 size is 111,619 bp, representing a seven fold improvement from the apple genome assembled by Velasco *et al*. (2010). Further annotation analyses predicted 53,922 protein-coding genes and 2,765 non-coding RNA genes.

A high-quality apple reference genome of *Malus domestica* 'Golden Delicious' doubled-haploid tree (GDDH13 v1.1) was assembled *de novo* and arranged into 17 chromosomes of apple using Illumina and long sequencing reads (PacBio), along with scaffolding based on optical maps (BioNano). The estimated genome size of GDDH13 v1.1 is 651Mb, from which 649.7 Mb (99.8%) was assembled. The N50 of contigs was much higher at 5.558 Mb. A total of 42,140 protein-coding genes (which represent 23.3% of the genome assembly) and 1,965 non-protein-coding genes were identified, however this was lower compared to the apple genome assembled by Velasco *et al*. 2010. The GDDH13 v1.1 currently serves as the reference genome for apple, and it was also utilised in this chapter (chapter 5) of this thesis.

Other *Rosaceae* crops sequenced genomes include: strawberry (Shulaev *et al.*, 2011); apricot (Zhang *et al.*, 2012), peach (Verde *et al.*, 2013), Chinese pear (Wu *et al.*, 2013), European pear (Chagné *et al.*, 2014). The comparison of the apple genomes to the other sequenced fruit crops in *Rosaceae* is summarised in Table 2.4.

**Table 2.4 Summary comparison of the apple genome sequences to other sequenced *Rosaceae'* fruit crops.**

| Fruit crop/cultivar name (species name) | Putative genes | Genome size (Mb) | Coverage (%) | Chromosome number | Contigs | Sequencing technology | References |
|---|---|---|---|---|---|---|---|
| Apple 'Golden Delicious' (*Malus x domestica*) | 57 386 | 603.9 | 81.3 | 2n=2x=34 | 122 146 | Whole-genome shotgun | Velasco *et al*., 2010 |
| Apple 'Golden Delicious' (*Malus x domestica*) | 53 922 | 632.4 | 90.0 | 2n=2x=34 | 111 619 | PacBio | Li *et al*., 2016 |
| Apple 'Golden Delicious' GDDH13 (*Malus x domestica*) | 42 140 | 649.7 | 99.8 | 2n=2x=34 | 5.558 Mb | Illumina, PacBio, BioNano | Daccord *et al*., 2017 |
| Woodland strawberry 'Hawaii 4' (*Fragaria vesca*) | 34 809 | 240.0 | 95.0 | 2n=2x=14 | 45 592 | Roche/454, Illumina/Solexa Life Technologies/SOLiD | Shulaev *et al*., 2011 |
| Japanese apricot (*Prunus mume*) | 31 390 | 280.0 | 84.6 | 2n=2x=14 | 45 592 | Whole-genome shotgun | Zhang *et al*., 2012 |
| Peach 'Lovell' (*Prunus persica*) | 27 852 | 265.0 | 81.5 | 2n=2x=16 | - | Sanger whole-genome shotgun | Verde *et al*., 2013 |
| Chinese pear 'Dangshansuli' (*Pyrus bretschneideri*) | 42 812 | 512.0 | 97.1 | 2n=2x=34 | 25 312 | BAC-by-BAC and next-generation sequencing | Wu *et al*., 2013 |
| European pear 'Bartlett' (*Pyrus communis*) | 43 419 | 577.3 | 96.2 | 2n=2x=34 | 182 196 | Roche 454 | Change *et al*., 2014 |

## 2.9 Transcriptomics and gene expression profiling in apple

Transcriptomics began in the mid 1990's as a field of study with the aim of profiling the transcriptome (Velculescu *et al.*, 1997; Piétu *et al.*, 1999). The term 'transcriptome' encompasses all the genomic counterparts which are expressed as RNA transcripts, including mRNA, rRNA, tRNA and other non-coding RNA molecules produced in a cell (Wang *et al.*, 2009). Transcriptome analysis is essential for the characterisation of developmental and functional elements of a genome, thus unravelling the complexity of genetic expression. Over the past decades, global analysis of gene expression has emerged as a powerful tool for elucidating expressed genes and biological pathways under different conditions (Haynes *et al.*, 2013; Chien *et al.*, 2015).

Several technologies have been developed in assessing gene expression, including the earlier traditional Northern blot (Alwine *et al.*, 1977), Sanger sequencing of cDNA and expressed sequence tag (EST) libraries (Adams *et al.*, 1991) as well as quantitative reverse transcription polymerase chain reaction (qRT-PCR) (Gibson *et al.*, 1996). Up to this point, sequencing was limited to gene discovery and RNA quantification due to high cost and low sequence depth. Shortly after, these techniques were surpassed by tag-based approaches which generate digital gene expression profiles in a high-throughput manner, such as serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995; Hu and Polyak, 2006), massively parallel signature sequencing (MPSS) (Brenner *et al.*, 2000; Peiffer *et al.*, 2008), and cap analysis of gene expression (CAGE) (Shiraki *et al.*, 2003; Kodzius *et al.*, 2006).

From the mid-1990s, DNA microarrays were the predominant method used for gene expression profiling (DeRisi *et al*., 1996). DNA microarrays are typically glass slides consisting of thousands of microscopic spots of DNA probes arranged in orderly rows and columns on a solid surface, with each spot corresponding to a different gene. The orderly arrangement of DNA probes makes it easy to identify a particular gene sequence according to the location of each spot. Two common types of microarray technologies exists amongst others including spotted cDNA microarrays which utilises cDNA probes spotted onto a glass slide robotic arrays (Schena *et al.*, 1995), and oligonucleotide arrays which employs an *in situ* synthesized oligonucleotides of various lengths (Lockhart *et al.*, 1996; Lipshutz *et al.*, 1999; Pozhitkov *et al.*, 2007). The use of DNA microarrays have provided

key insight towards understanding the phenomenon behind the genetics and regulation of gene expression.

Gene expression studies in apple using microarrays have been undertaken in elucidating the molecular mechanisms of early fruit development (Lee *et al.*, 2007; Janssen *et al.*, 2008), aroma production and red coloration (Schaffer *et al.*, 2007), and fruit development (Janssen *et al.*, 2008; Soria-Guerra *et al.*, 2011). Gene expression analysis in apple has also been undertaken by the generation of ESTs (Newcomb *et al.*, 2006; Wisniewski *et al.*, 2008), cDNA-SSH (Norelli *et al.*, 2009), cDNA-AFLP (Baldo *et al.*, 2010) and microarrays (Soglio *et al.*, 2009; Sarowar *et al.*, 2011). Nowadays, DNA microarrays are a relatively inexpensive and mature technology. Nonetheless, this technology relies on nucleic-acid hybridisation and presents inherent limitation accompanied by high levels of background noise resulting from cross-hybridisation artefacts, thus compromising the accuracy of the expression measurements. The use of microarrays also limits the number of interrogated transcripts, particularly transcripts present in low abundance (Gautier *et al.*, 2004; Mortazavi *et al.*, 2008; Wang *et al.*, 2009). Therefore, DNA microarrays are not necessarily well suited for measuring absolute expression levels (Palmieri and Schlötterer, 2009). The next-generation sequencing (NGS) technologies have since revolutionised transcriptomic studies with the advent of RNA-sequencing (RNA-seq) (Morozova *et al.*, 2009; Wang *et al.*, 2009).

### 2.9.1 RNA sequencing and differential gene expression

RNA sequencing (RNA-seq) is the first sequencing-based method that allows the entire transcriptome to be surveyed simultaneously at base pair resolution, and over a higher dynamic range of expression levels without the requirement of a *priori* knowledge of transcribed regions (Morozova *et al.*, 2009; Wang *et al.*, 2009). RNA-seq is an application of next-generation sequencing methods to study, quantify and analyse the transcriptome for transcriptional changes between different conditions. One of the most common applications of RNA-Seq is to investigate differential gene expression (DGE), where up-regulated and down-regulated genes can be identified (Soneson *et al.*, 2016). This technology is also used in detecting alternative splicing (Gan *et al.*, 2010), detecting gene fusion events (Vu *et al.*, 2018), identifying novel transcripts (Steijger *et al.*, 2013), and identifying splice junctions (Jambagi and Dunwell, 2015).

The primary purpose of RNA sequencing (RNA-Seq) is to gain insight into the biological processes, cellular components, and molecular function of putative gene products involved. Gene products that are differentially expressed can be annotated with BLAST and Blast2GO to assign Gene Ontology terms (Conesa and Götz, 2008; Götz *et al.*, 2008, 2011). Prediction and visualisation of the metabolic pathways within the transcriptome can be undertaken in Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2006; Hattori *et al.*, 2008; Götz *et al.*, 2008).

RNA sequencing analysis starts with RNA extraction from biological samples, followed by library preparation, where extracted RNA is reverse transcribed into double stranded cDNA. The cDNA reads are attached to universal adapter sequences and DNA barcodes are added to one or both ends of each fragment (single or paired-end sequencing reads, respectively). Thereafter sequencing is performed on a NGS platform (e.g. Illumina HiSeq), producing millions of so-called reads. However, the raw datasets produced are large and complex. As with any high-throughput technology, large datasets necessitate computer aided sequence and statistical analysis (Oshlack *et al.*, 2010; Kukurba and Montgomery, 2015). The resulting reads are either aligned to a reference genome or transcriptome. The reads can be assembled as a *de novo* transcriptome if the reference genome or transcriptome is not available. The main aim is to identify the genomic location where each short read best matches to the reference genome. The mapped reads are then assembled into gene-level, exon-level or transcript-level summaries. The summarized data are then normalised for gene length and total transcript, where the expression differences are compared through statistical tests, leading to a ranked list of genes with associated *p*-values, i.e. discovery of differential expressed genes (Wang *et al.*, 2009; Trapnell *et al.*, 2010; Conesa *et al.*, 2016). An overview of a typical RNA-seq analysis workflow is also outlined in Figure 2.4 (Oshlack *et al.*, 2010).

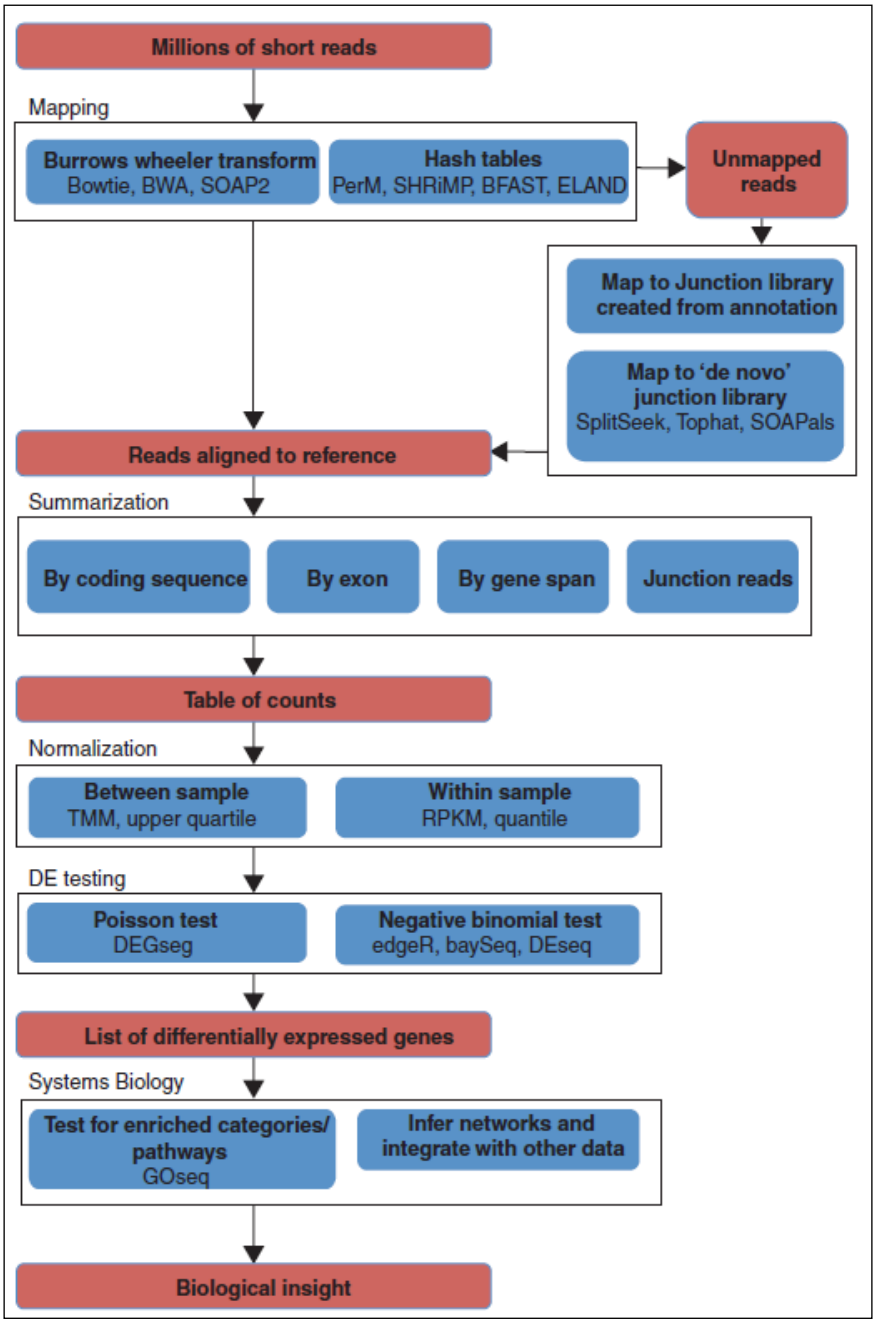


**Figure 2.4 Typical workflow demonstrating an overview of RNA sequencing (RNA-seq) data analysis**. The analysis steps typically consists of five steps highlighted in red boxes: Mapping of reads, reads alignment to reference, table of counts (normalisation), differential gene expression analysis, and biological annotations (pathways). The methodological and software examples are shown in blue boxes. (Adapted from: Oshlack *et al.*, 2010).

The choice of aligneent software depends mainly on the nature of the species and sequencing platform (Conesa *et al.*, 2016). Currently, various software packages for read alignment exist but the two most widely used are Hash Look-up Table Algorithms (Hash-Table) and Burrows-Wheeler Transform (BWT)-based mapping alignement methods. The Hash-Table includes GSNAP (Wu and Nacu, 2010) and SOAP (Li *et al.*, 2008). The BWT-based mapping aligners are Bowtie (Langmead *et al.*, 2009; Langmead, 2013), BWA (Li and Durbin, 2010), amongst others. For *de novo* assembly, a number of *de novo* transcriptome assemblers are currently available, such as Velvet (Zerbino and Birney, 2008), ABySS (Birol *et al.*, 2009), SOAPdenovo (Li *et al.*, 2008, 2010; Xie *et al.*, 2014), and Trinity (Grabherr *et al.*, 2011), all assemble the sequence reads directly into transcripts.

One of the main goals of differential expression analysis is to identify genes which are differentially expressed between two or more conditions. Such genes are selected based on a combination of expression level threshold and expression score cutoff, which is based on *p*-values generated by statistical modelling (Conesa *et al.*, 2016). The expression level of each RNA unit is measured by counting the number of reads per gene i.e. the number of sequenced fragments that map to the transcript, using two alternative metrices. The first is,"reads" per kilobase per million mapped reads (RPKM) and the sec is, "fragment" per kilobase per million mapped reads (FPKM). The metrices arethe number of reads or fragments generated from each transcript based on the length and depth of the sequencing. Additionally, they are usedas a measure of the estimated expression level of each transcript based on the length (Mortazavi *et al.*, 2008; Wagner *et al.*, 2012; Cai *et al.*, 2018). Nonetheless, RPKM and FPKM complicates normalisation (Bullard *et al.*, 2010). Different computational pipelines such as Cufffdiff and Cufflinks (Trapnell *et al.*, 2010), DEGseq (Wang *et al.*, 2010), DESeq (Anders and Huber, 2010; Wu *et al.*, 2012; Love *et al.*, 2014), edgeR (Robinson *et al.*, 2010), GFOLD (Feng *et al.*, 2012) and NOISeq (Tarazona *et al.*, 2011; Ding *et al.*, 2015) have been designed to normalise measurements across samples and accurately detect gene transcripts that change in expression levels.

The NGS RNA-seq platforms have been utilised extensively in gene expression profiling of columnar growth habit and hormonal state of shoot apical meristem in apple (Krost *et al.*, 2012, 2013; Li *et al.*, 2012). Characterisation of ontogenic scab resistance in apple (Gusberti *et al.*, 2013), exploration of changes associated with internal browning of apple during postharvest storage (Mellidou *et al.*, 2014). RNA-seq system has been utilised in apple studies to identify genes related

to fruit acidity (Bai *et al.*, 2015), anthocyanin deficient yellow skin (El-Sharkawy *et al.*, 2015), fruitlet abscission in apple seeds (Ferrero *et al.*, 2015), and changes in apple peel tissues during $CO_2$ injury symptom development (Johnson and Zhu, 2015). Thereon, numerous transcriptomic studies has been conducted including: transcription profiling of the chilling requirement for bud break in apples (Porto *et al.*, 2015), analysis of apple leaves in response to infection *Alternaria* blotch disease (Zhu *et al.*, 2017), transcriptomic responses to biotic stresses related to fungi, virus and bacteria attacks (Balan *et al.*, 2018) and transcriptomic analysis of red-fleshed apples (Wang *et al.*, 2018).

## 2.10 References

**Adams M, M Kelley J, D Gocayne J, Dubnick M, Polymeropoulos M, Xiao H, Merril C, Wu A, Olde B, Moreno R**. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. Science **252**, 1651–1656.

**Alston FH**. 1976. Dwarfing and lethal genes in apple progenies. Euphytica **25**, 505–514.

**Alwine JC, Kemp DJ, Stark GR**. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences of the United States of America **74**, 5350–5354.

**Anders S, Huber W**. 2010. Differential expression analysis for sequence count data. Genome Biology **11**, R106.

**Anderson MA, Cornish EC, Mau S-L, *et al.*** 1986. Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature **321**, 38–44.

**Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troggio M, Sargent DJ**. 2012. Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* infinium whole genome genotyping array. BMC Genomics **13**, 203.

**Appleby N, Edwards D, Batley J**. 2009. New technologies for ultra-high throughput genotyping in plants. Methods in Molecular Biology **513**, 19–39.

**Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A**. 1999. Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the α-subunit of GTP-binding protein. Proceedings of the National Academy of Sciences of the United States of America **96**, 10284–10289.

**Bai Y, Dougherty L, Cheng L, Zhong GY, Xu K**. 2015. Uncovering co-expression gene network

modules regulating fruit acidity in diverse apples. BMC Genomics **16**, 612.

**Balan B, Marra FP, Caruso T, Martinelli F**. 2018. Transcriptomic responses to biotic stresses in *Malus* x *domestica*: A meta-analysis study. Scientific Reports **8**, 1970.

**Baldo A, Norelli JL, Farrell RE, Bassett CL, Aldwinckle HS, Malnoy M**. 2010. Identification of genes differentially expressed during interaction of resistant and susceptible apple cultivars (*Malus* × *domestica*) with *Erwinia amylovora*. BMC Plant Biology **10**, 1–10.

**Bassil N V., Davis TM, Zhang H, *et al.*** 2015. Development and preliminary evaluation of a 90K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria* × *ananassa*. BMC Genomics **16**, 1–30.

**Bateman AJ**. 1955. Self-incompatibility systems in angiosperms: III. *Cruciferae*. Heredity **9**, 53–68.

**Bianco L, Cestaro A, Linsmith G, *et al.*** 2016. Development and validation of the Axiom® apple 480K SNP genotyping array. Plant Journal **86**, 62–74.

**Bianco L, Cestaro A, Sargent DJ, *et al.*** 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh.). PLoS One **9**, e110377.

**Birol I, Jackman SD, Nielsen CB, *et al.*** 2009. De novo transcriptome assembly with ABySS. Bioinformatics **25**, 2872–2877.

**Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D**. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biology **5**, 1962–1972.

**Bomblies K, Weigel D**. 2007. Hybrid necrosis: Autoimmunity as a potential gene-flow barrier in plant species. Nature Reviews Genetics **8**, 382–393.

**Bošković R, Tobutt KR**. 1996. Correlation of stylar ribonuclease zymograms with incompatibility alleles in sweet cherry. Euphytica **90**, 245–250.

**Brenner S, Johnson M, Bridgham J, *et al.*** 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotechnology **18**, 630–634.

**Brookes AJ**. 1999. The essence of SNPs. Gene **234**, 177–186.

**Broothaerts W**. 2003. New findings in apple S-genotype analysis resolve previous confusion and request the re-numbering of some S-alleles. Theoretical and Applied Genetics **106**, 703–714.

**Broothaerts W, Janssens GA, Proost P, Broekaert WF**. 1995. cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. Plant Molecular Biology **27**, 499–511.

**Brown S**. 1992. Genetics of apple. Plant Breeding Reviews **9**, 333–366.

**Bullard JH, Purdom E, Hansen KD, Dudoit S**. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics **11**, 94.

**Bus VGM, Chagné D, Bassett HCM, *et al.*** 2008. Genome mapping of three major resistance genes to woolly apple aphid (*Eriosoma lanigerum* Hausm.). Tree Genetics and Genomes **4**, 223–236.

**Bus V, Ranatunga C, Gardiner S, Bassett H, Rikkerink E, Geibel M, Fischer M, Fischer C**. 2000. Marker assisted selection for pest and disease resistance in the New Zealand apple breeding programme. Acta Horticulturae **538**, 541–547.

**Cai G, Franks JM, Whitfield ML**. 2018. RNA-seq analyses of molecular abundance (RoMA) for detecting differential gene expression. https://doi.org/10.1101/410985.

**Campbell CS, Evans RC, Morgan DR, Dickinson TA, Arsenault MP**. 2007. Phylogeny of subtribe *Pyrinae* (formerly the *Maloideae*, *Rosaceae*): Limited resolution of a complex evolutionary history. Plant Systematics and Evolution **266**, 119–145.

**Cane J, Schiffhauer D**. 2003. Dose-response relationships between pollination and fruiting refine pollinator comparisons for cranberry (*Vaccinium macrocarpon* [*Ericaceae*]). American Journal of Botany **90**, 1425–1432.

**Celton JM, Tustin DS, Chagné D, Gardiner SE**. 2009. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. Tree Genetics and Genomes **5**, 93–107.

**Chagné D, Batley J, Edwards D, Forster JW**. 2007. Single nucleotide polymorphism genotyping in plants. In: Oraguzie N.C., Rikkerink E.H.A., Gardiner S.E., De Silva H.N. (Eds). Association Mapping in Plants. Springer, New York, 77–94.

**Chagné D, Crowhurst RN, Pindo M, *et al.*** 2014. The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). PLoS One **9**, e92644.

**Chagné D, Crowhurst RN, Troggio M, *et al.*** 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One **7**, e31745.

**Chandra K, Pandey A**. 2017. QTL mapping in crop improvement: A basic concept. International Journal of Current Microbiology and Applied Sciences **6**, 835–842.

**Chen C, Zhiguo E, Lin HX**. 2016. Evolution and molecular control of hybrid incompatibility in plants. Frontiers in Plant Science **7**, 1–10.

**Chien CH, Chow CN, Wu NY, Chiang-Hsieh YF, Hou PF, Chang WC**. 2015. EXPath: A database of comparative expression analysis inferring metabolic pathways for plants. BMC Genomics **16**, S6.

**Clark MD, Schmitz CA, Rosyara UR, Luby JJ, Bradeen JM**. 2014. A consensus 'Honeycrisp' apple (*Malus* × *domestica*) genetic linkage map from three full-sib progeny populations. Tree Genetics and Genomes **10**, 627–639.

**Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK**. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica **142**, 169–196.

**Collard BCY, Mackill DJ**. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **363**, 557–572.

**Conesa A, Götz S**. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics, 1–13.

**Conesa A, Madrigal P, Tarazona S, *et al.*** 2016. A survey of best practices for RNA-seq data analysis. Genome Biology **17**, 1–19.

**Cornille A, Gladieux P, Giraud T**. 2013. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. Evolutionary Applications **6**, 737–748.

**Cornille A, Gladieux P, Smulders MJM, *et al.*** 2012. New insight into the history of domesticated apple: Secary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genetics **8**, e1002703.

**Costes E, García-Villanueva E**. 2007. Clarifying the effects of dwarfing rootstock on vegetative and reproductive growth during tree development: A study on apple trees. Annals of Botany **100**, 347–357.

**DAFF**. 2018. A profile of the South African apple market value chain. Pretoria: Directorate Marketing. South African Department of Agriculture Fisheries and Forestry. http://www.daff.gov.za.

**Daviere J-M, Achard P**. 2013. Gibberellin signaling in plants. Development **140**, 1147–1151.

**Dennis FG**. 1994. Dormancy: What we know (and don't know). HortScience **29**, 1249–1255.

**Deschamps S, Llaca V, May GD**. 2012. Genotyping-by-sequencing in plants. Biology **1**, 460–483.

**Dill A, Jung HS, Sun TP**. 2001. The DELLA motif is essential for gibberellin-induced degradation of RGA. Proceedings of the National Academy of Sciences of the United States of America **98**, 14162–14167.

**Ding B, Zheng L, Zhu Y, Li N, Jia H, Ai R, Wildberg A, Wang W**. 2015. Normalization and noise reduction for single cell RNA-seq experiments. Bioinformatics **31**, 2225–2227.

**Edwards D, Forster JW, Chagné D, Batley J**. 2007. What are SNPs? *In*: Oraguzie NC, In: Rikkerink EHA, Gardiner SE, de Silva HN, (Eds). *Association Mapping in Plants*. Springer-Verlag New York, 41–52.

**El-Sharkawy I, Liang D, Xu K**. 2015. Transcriptome analysis of an apple (*Malus* × *domestica*) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. Journal of Experimental Botany **66**, 7359–7376.

**Ellegren H**. 2004. Microsatellites: Simple sequences with complex evolution. Nature Reviews Genetics **5**, 435–445.

**Estoup A, Jarne P, JM C**. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology **11**, 1591–1604.

**Falginella L, Cipriani G, Monte C, Gregori R, Testolin R, Velasco R, Troggio M, Tartarini S**. 2015. A major QTL controlling apple skin russeting maps on the linkage group 12 of 'Renetta Grigia di Torriana'. BMC Plant Biology **15**, 1–13.

**Fallahi E, Colt WM, Fallahi B, Chun I**. 2002. The importance of apple rootstocks on tree growth, yield, fruit quality, leaf nutrition, and photosynthesis with an emphasis on 'Fuji'. HortTechnology **12**, 38–44.

**Farneti B, Guardo M Di, Khomenko I, Cappellin L, Biasioli F, Velasco R, Costa F**. 2017. Genome-wide association study unravels the genetic control of the apple volatilome and its interplay with fruit texture. Journal of Experimental Botany, **68**, 1467–1478.

**Feng J, Meyer CA, Liu JS, Wang Q, Shirley Liu X, Zhang Y**. 2012. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics **28**, 2782–2788.

**Fernández-Fernández F, Evans KM, Clarke JB, Govan CL, James CM, Marič S, Tobutt KR**. 2008. Development of an STS map of an interspecific progeny of *Malus*. Tree Genetics and Genomes **4**, 469–479.

**Fernández-Fernández F, Padmarasu S, Šurbanovski N, Evans KM, Tobutt KR, Sargent DJ**.

2014. Characterisation of the virescent locus controlling a recessive phenotype in apple rootstocks (*Malus pumila* Mill.). Molecular Breeding **33**, 373–383.

**Ferrero S, Carretero-Paulet L, Mendes MA, Botton A, Eccher G, Masiero S, Colombo L**. 2015. Transcriptomic signatures in seeds of apple (*Malus domestica* L. Borkh) during fruitlet abscission. PLoS One **10**, e0120503.

**Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC**. 2003. Collection, maintenance, characterization, and utilization of wild apples of Central Asia. *In*: Janick J, (Eds). *Horticultural Reviews*. John Wiley & Sons, Inc. Vol 29.1–62.

**Fridborg I, Kuusk S, Robertson M, Sundberg E**. 2001. The *Arabidopsis* protein SHI represses gibberellin responses in *Arabidopsis* and barley. Plant Physiology **127**, 937–948.

**Fujioka S, Yokota T**. 2003. Biosynthesis and metabolism of brassinosteroids. Annual Review of Plant Biology **54**, 137–164.

**Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X**. 2010. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. Cell Research **20**, 763–783.

**Ganal MW, Altmann T, Röder MS**. 2009. SNP identification in crop plants. Current Opinion in Plant Biology **12**, 211–217.

**Ganal MW, Polley A, Graner EM, Plieske J, Wieseke R, Luerssen H, Durstewitz G**. 2012. Large SNP arrays for genotyping in crop plants. Journal of Biosciences **37**, 821–828.

**Gao ZS, van de Weg WE**. 2006. The *Vf* gene for scab resistance in apple is linked to sub-lethal genes. Euphytica **151**, 123–132.

**Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA**. 2007. Apple. *In*: Kole C, (Eds). *Genome Mapping* and *Molecular Breeding* in *Plants***.** Springer-Verlag Berlin Heidelberg, pp. 1–62.

**Gautier L, Møller M, Friis-Hansen L, Knudsen S**. 2004. Alternative mapping of probes to genes for Affymetrix chips. BMC Bioinformatics **5**, 111.

**Germain-Aubrey CC, Nelson C, Soltis DE, Soltis PS, Gitzendanner MA**. 2016. Are microsatellite fragment lengths useful for population-level studies? The case of *Polygala lewtonii* (*Polygalaceae*). Applications in Plant Sciences **4**, 1500115.

**Gessler C, Patocchi A**. 2007. Recombinant DNA technology in apple. Advances in Biochemical Engineering, Biotechnology, 107: 113-132.

**Ghosh S, Halder S**. 2018. Effect of different kinds of gibberellin on temperate fruit crops : A review. The Pharma Innovation Journal **7**, 315–319.

**Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C**. 1998. Simple sequence repeats for the genetic analysis of apple. Theoretical and Applied Genetics **96**, 1069–1076.

**Gibbs P**. 1988. Self-incompatibility mechanism in flowering plants: Some complications and clarification. Lagascalia **15**, 17–28.

**Gibson UEM, Heid CA, Williams PM**. 1996. A novel method for real time quantitative RT-PCR. Genome Research **6**, 995–1001.

**González-Robles A, Manzaneda AJ, Bastida JM, Harvey N, Jaime R, Salido T, Martínez LM, Fernández-Ocaña A, Alcántara JM, Rey PJ**. 2016. Development and characterization of microsatellite primers in the endangered Mediterranean shrub *Ziziphus lotus* (*Rhamnaceae*). Applications in Plant Sciences **4**, 1600092.

**Götz S, Arnold R, Sebastian-Leon P, Martin-Rodriguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T, Conesa A**. 2011. B2G-FAR, a species-centered GO annotation repository. Bioinformatics **27**, 919–924.

**Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A**. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research **36**, 3420–3435.

**Grabherr MG, Haas BJ, Yassour M, *et al.*** 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology **29**, 644–652.

**Grattapaglia D, Sederoff R**. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics **137**, 1121–1137.

**Gross L**. 2007. Autoimmunity: A barrier to gene flow in plants? PLoS Biology **5**, e262.

**Di Guardo M, Bink MCAM, Guerra W, *et al.*** 2017. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. Journal of Experimental Botany **68**, 1451–1466.

**Gupta PK, Roy JK, Prasad M**. 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science **80**, 524–535.

**Gusberti M, Gessler C, Broggini GAL**. 2013. RNA-Seq analysis reveals candidate genes for ontogenic resistance in *Malus-Venturia* Pathosystem. PLoS One **8**, e78457.

**Haldane JBS**. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. Journal of Genetics **8**, 299–309.

**Han Y, Zheng D, Vimolmangkang S, Khan MA, Beever JE, Korban SS**. 2011. Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. Journal of Experimental Botany **62**, 5117–5130.

**Hancock J, Luby J, Brown S, Lobos G**. 2008. Apples. *In*: Hancock JF, (Eds). *Temperate Fruit Crop Breeding*. Springer, Dordrecht. pp. 1–38.

**Harris SA, Robinson JP, Juniper BE**. 2002. Genetic clues to the origin of the apple. Trends in Genetics **18**, 426–430.

**Hattori M, Itoh M, Araki M, *et al.*** 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research **36**, 480–484.

**Haynes WA, Higdon R, Stanberry L, Collins D, Kolker E**. 2013. Differential expression analysis for pathways. PLoS Computational Biology **9**, e1002967.

**Hedden P, Thomas S**. 2012. Gibberellin biosynthesis and its regulation. The Biochemical Journal **444**, 11–25.

**HORTGRO**. 2018. Key Deciduous Fruit Statistics: Pome Fruit. http://www.hortgro.co.za.

**Hosseinzadeh-Colagar A, Haghighatnia MJ, Amiri Z, Mohadjerani M, Tafrihi M**. 2016. Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. Molecular Biology Research communications **5**, 167–174.

**Hou S, Liu Z, Shen H, Wu D**. 2019. Damage-associated molecular pattern-triggered immunity in plants. Frontiers in Plant Science **10**, 646.

**Houston RD, Taggart JB, Cézard T, *et al.*** 2014. Development and validation of a high density SNP genotyping array for *Atlantic salmon* (*Salmo salar*). BMC Genomics **15**, 90.

**Howard NP, van de Weg E, Bedford DS, Peace CP, Vanderzande S, Clark MD, Teh SL, Cai L, Luby JJ**. 2017. Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus* × *domestica*) pedigree-connected SNP data set. Horticulture Research **4**, 17003.

**Hu M, Polyak K**. 2006. Serial analysis of gene expression. Nature Protocol **1**, 1743–1760.

**Hummer KE, Janick J**. 2009. *Rosaceae*: Taxonomy, economic importance, genomics. *In*: Folta

KM, Gardiner SE, (Eds). *Genetics and Genomics of Rosaceae*. Springer, New York, USA. pp. 1–17.

**Huq A, Akter S, Sup I, Hoy N, Kim T, Jin Y, Kwon J, Kang K**. 2016. Identification of functional SNPs in genes and their effects on plant phenotypes. Journal of Plant Biotechnology **43**, 1–11.

**Jackson JE**. 2003. The growing of apple and pears. *In*: Biology of Apples and Pears. Cambridge University Press, New York, pp. 4-20.

**Jambagi S, Dunwell JM**. 2015. Global transcriptome analysis and identification of differentially expressed genes after infection of *Fragaria vesca* with powdery mildew (*Podosphaera aphanis*). Transcriptomics **3**, 1000106.

**Janick J, Cummins J, Brown S, Hemmat M**. 1996. Apples. *In*: Janick J, Moore JN, (Eds). *Fruit Breeding, Tree and Tropical Fruits*. John Wiley & Sons, Inc., Vol I.1–77.

**Jannink JL, Lorenz AJ, Iwata H**. 2010. Genomic selection in plant breeding: From theory to practice. Briefings in Functional Genomics and Proteomics **9**, 166–177.

**Jansen J, De Jong AG, Van Ooijen JW**. 2001. Constructing dense genetic linkage maps. Theoretical and Applied Genetics **102**, 1113–1122.

**Janssen BJ, Thodey K, Schaffer RJ, *et al.*** 2008. Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. BMC Plant Biology **29**, 1–29.

**Janssens GA, Goderis IJ, Broekaert WF, Broothaerts W**. 1995. A molecular method for S-allele identification in apple based on allele-specific PCR. Theoretical and Applied Genetics **91**, 691–698.

**Jehan T, Lakhanpaul S**. 2006. Single nucleotide polymorphism (SNP) - methods and applications in plant genetics: A review. Indian Journal of Biotechnology **5**, 435–459.

**Johnson FT, Zhu Y**. 2015. Transcriptome changes in apple peel tissues during $CO_2$ injury symptom development under controlled atmosphere storage regimens. Horticulture Research **2**, 15061.

**Jones J, Dangl J**. 2006. The plant immune system. Nature **444**, 323–329.

**Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D**. 2008. GDR (Genome Database for *Rosaceae*): Integrated web-database for *Rosaceae* genomics and genetics data. Nucleic Acids Research **36**, 1034–1040.

**Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Masumi I, Kawashima S, Katayama T,**

**Araki M, Mika H**. 2006. From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Research **34**, D354–D357.

**Kao T, Huang S**. 1994. Gametophytic self-incompatibility : A mechanism for self/nonself discrimination during sexual reproduction. Plant Physiology **105**, 461–466.

**Kao TH, McCubbin AG**. 1996. How flowering plants discriminate between self and non-self pollen to prevent inbreeding. Proceedings of the National Academy of Sciences **93**, 12059–12065.

**Kao T, Tsukamoto T**. 2004. The molecular and genetic bases of *S-RNase*-based self-incompatability. The Plant Cell **16**, S72–S83.

**Karuppanapandian T, Moon JC, Kim C, Manoharan K, Kim W**. 2011. Reactive oxygen species in plants: Their generation, signal transduction, and scavenging mechanisms. Australian Journal of Crop Science **5**, 709–725.

**Kim HT, Moriya S, Okada K, Abe K, Park JI, Yamamoto T, Nou S**. 2016. Identification and characterization of S-*RNase* genes in apple rootstock and the diversity of *S-RNases* in *Malus* species. Journal of Plant Biotechnology **43**, 49–57.

**Kobel F**. 1939. Weitere Untersuchungen uber die Befruchtungsverhaltnisse der Apfelund Birnsorten. Landw Jahrb Schweiz **53**, 160–191.

**Kodzius R, Kojima M, Nishiyori H, *et al.*** 2006. CAGE: cap analysis of gene expression. Nature Methods **3**, 211–222.

**Komori S, Soejima J, Abe K, Kotoda N, Ito Y, Bessho H**. 2000. Reanalysis of the *S*-allele genotypes in several apple cultivars. Journal of the Japanese Society for Horticultural Science **69**, 449–459.

**Korban SS, Skirvin RM**. 1984. Nomenclature of the cultivated apple. HortScience **19**, 177–180.

**Kosambi DD**. 1944. The estimation of map distances from recombination values. Annals of Eugenics **12**, 172–175.

**Krost C, Petersen R, Lokan S, Schmidt ER**. 2013. Evaluation of the hormonal state of columnar apple trees (*Malus* x *domestica* ) based on high throughput gene expression studies. Plant Molecular Biology **81**, 211–220.

**Krost C, Petersen R, Schmidt ER**. 2012. The transcriptomes of columnar and standard type apple trees (*Malus* x *domestica* ) — A comparative study. Gene **498**, 223–230.

**Kukurba KR, Montgomery SB**. 2015. RNA sequencing and analysis. Cold Spring Harbor

Protocols **11**, 951–969.

**Kumar S, C. A. M. Bink M, K. Volz R, Bus V, Chagne D**. 2011. Towards genomic selection in apple (*Malus* × *domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. Tree Genetics & Genomes **8**, 1–14.

**Langmead**. 2013. Fast gapped-read alignment with Bowtie 2. Nature Methods **9**, 357–359.

**Langmead B, Trapnell C, Pop M, Salzberg SL**. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10**, R25.

**Leclercq SB, Rivals E, Jarne P**. 2010. DNA slippage occurs at microsatellite loci without minimal threshold length in humans: A comparative genomic approach. Genome Biology and Evolution **2**, 325–335.

**Lee YP, Yu GH, Seo YS, Han SE, Choi YO, Kim D, Mok IG, Kim WT, Sung SK**. 2007. Microarray analysis of apple gene expression engaged in early fruit development. Plant Cell Reports **26**, 917–926.

**Leforestier D, Ravon E, Muranty H, Cornille A, Lemaire C, Giraud T, Durel CE, Branca A**. 2015. Genomic basis of the differences between cider and dessert apple varieties. Evolutionary Applications **8**, 650–661.

**Levinson G, Gutman GA**. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Molecular Biology and Evolution **4**, 203–221.

**Lewak S**. 2011. Metabolic control of embryonic dormancy in apple seed: Seven decades of research. Acta Physiologiae Plantarum **33**, 1–24.

**Li H, Durbin R**. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**, 589–595.

**Li X, Kui L, Zhang J, *et al.*** 2016. Improved hybrid *de novo* genome assembly of domesticated apple (*Malus* x *domestica*). GigaScience **5**, 35.

**Li R, Li Y, Kristiansen K, Wang J**. 2008. SOAP: Short oligonucleotide alignment program. Bioinformatics **24**, 713–714.

**Li T, Long S, Li M, Bai S, Zhang W**. 2012. Determination of *S*-genotypes and identification of five novel *S-RNase* alleles in wild *Malus* species. Plant Molecular Biology Reporter **30**, 453–461.

**Li X, Singh J, Qin M, Li S, Zhang X, Zhang M, Khan A, Zhang S, Wu J**. 2019. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly

improvement and GWAS in pear (*Pyrus*). Plant Biotechnology Journal, https://doi.org/10.1111/pbi.13085.

**Li D, Wang S, Shen Y, Meng X, Xu X, Wang R, Li J**. 2018. A multiplex microsatellite PCR method for evaluating genetic diversity in grass carp (*Ctenopharyngodon idellus*). Aquaculture and Fisheries **3**, 238–245.

**Li R, Zhu H, Ruan J, *et al.*** 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Research **20**, 265–272.

**Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C**. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus* x *domestica* Borkh.). Molecular Breeding **10**, 217–241.

**Lipshutz R, P.A. Fodor S, Gingeras T, J. Lockhart D**. 1999. High density synthetic oligonucleotide arrays. Nature Genetics **21**, 20–24.

**Lockhart DJ, Dong H, Byrne MC, *et al.*** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology **14**, 1675–1680.

**Long S, Li M, Han Z, Wang K, Li T**. 2010. Characterization of three new *S*-alleles and development of an S-allele-specific PCR system for rapidly identifying the *S*-genotype in apple cultivars. Tree Genetics and Genomes **6**, 161–168.

**Love MI, Huber W, Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**, 550.

**Mabberley DJ, Jarvis CE, Juniper BE**. 2001. The name of the apple. Telopea **9**, 421–430.

**Macho AP, Zipfel C**. 2014. Plant PRRs and the activation of innate immune signaling. Molecular Cell **54**, 263–272.

**Mahajan R, Gupta P**. 2012. Molecular markers: Their use in tree improvement. Journal of Forest Science **58**, 137–144.

**Maliepaard C, Alston FH, van Arkel G, *et al.*** 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theoretical and Applied Genetics **97**, 60–73.

**Matsumoto S**. 2014. Apple pollination biology for stable and novel fruit production: Search system for apple cultivar combination showing incompatibility, semicompatibility, and full-compatibility based on the *S-RNase* allele database. International Journal of Agronomy, https://doi.org/10.1155/2014/138271.

**Matsumoto D, Tao R**. 2016. Distinct self-recognition in the *Prunus S-RNase*-based gametophytic self-incompatibility system. The Horticulture Journal **85**, 289–305.

**McClure B**. 2009. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for. Journal of Experimental Botany, **60**, 1069–1081.

**McClure B, Franklin-Tong V**. 2006. Gametophytic self-incompatibility: Understanding the cellular mechanisms involved in " self " pollen tube inhibition. Planta **224**, 233–245.

**McClure BA, Gray JE, Anderson MA, Clarke AE**. 1989. Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature **347**, 955–957.

**Mellidou I, Buts K, Hatoum D, *et al.*** 2014. Transcriptomic events associated with internal browning of apple during postharvest storage. BMC Plant Biology **14**, 328.

**Mhamdi A, Van Breusegem F**. 2018. Reactive oxygen species in plant development. Development **145,** dev164376.

**Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam NK, Latif MA**. 2013. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. International Journal of Molecular Sciences **14**, 22499–22528.

**Minamikawa M, Kakui H, Wang S, Kotoda N, Kikuchi S, Koba T, Sassa H**. 2010. Apple *S* locus region represents a large cluster of related, polymorphic and pollen-specific F-box genes. Plant Molecular Biology **74**, 143–154.

**Mittler R, Vanderauwera S, Gollery M, Breusegem F Van**. 2004. Reactive oxygen gene network of plants. Trends in Plant Science **9**, 490–498.

**Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Breusegem F Van**. 2011. ROS signaling : The new wave ? Trends in Plant Science **16**, 300–309.

**Moniruzzaman M, Khatun R, Yaakob Z, Khan MS, Mintoo AA**. 2016. Development of Microsatellites: A powerful genetic marker. The Agriculturists **13**, 152–172.

**Montanari S, Brewer L, Lamberts R, Velasco R, Malnoy M, Perchepied L, Guérif P, Durel CE, Gardiner SE, Chagné D**. 2016. Genome mapping of postzygotic hybrid necrosis in an interspecific pear population. Horticulture Research **3**, 15064.

**Montanari S, Saeed M, Knäbel M, *et al.*** 2013. Identification of *Pyrus* single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific *Pyrus* hybrids. PLoS One **8**, e77022.

**Morgan J, Richards A**. 1993. The book of apples. Nature **366**, 641.

**Morozova O, Hirst M, Marra M**. 2009. Applications of new sequencing technologies for transcriptome analysis. Annual Review of Genomics and Human Genetics **10**, 135–151.

**Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B**. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods **5**, 621.

**Morton NE**. 1996. Logarithm of odds (lods) linkage in complex inheritance. Proceedings of the National Academy of Sciences of the United States of America **93**, 3471–3476.

**Nadeem MA, Nawaz MA, Shahid MQ, *et al.*** 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment **32**, 261–285.

**de Nettancourt D**. 1977. Incompatibility in angiosperms. Springer-Verlag Berlin Heidelberg, 1–2.

**Newcomb RD, Crowhurst RN, Gleave AP, *et al.*** 2006. Analyses of expressed sequence tags from apple. Plant Physiology **141**, 147–166.

**Norelli JL, Farrell RE, Bassett CL, Baldo AM, Lalli DA, Aldwinckle HS, Wisniewski ME**. 2009. Rapid transcriptional response of apple to fire blight disease revealed by cDNA suppression subtractive hybridization analysis. Tree Genetics and Genomes **5**, 27–40.

**Nyholt DR**. 2000. All LODs are not created equal. American Journal of Human Genetics **67**, 282–288.

**O'Brien JA, Daudi A, Butt VS, Bolwell GP**. 2012. Reactive oxygen species and their role in plant defence and cell wall metabolism. Planta **236**, 765–779.

**Okada K**. 2015. DNA markers and the molecular mechanism of self-(in)compatibility in Japanese pear (*Pyrus pyrifolia* Nakai). The Horticulture Journal **84**, 183–194.

**Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Lúcia M, Vieira C**. 2006. Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology **29**, 294–307.

**Van Ooijen JW**. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Research **93**, 343–349.

**Van Ooijen JW, Voorrips RE**. 2006. JoinMap 4: Software for the calculation of genetic linkage maps in experimental population. Kyazma B.V., Wageningen, Netherlands.

**Van Ooijen J, Voorrips RE**. 2017. JoinMap5: Software for the calculation of genetic linkage maps

in experimental populations of diploid species. Kyazma. B.V., Wageningen, Netherlands.

**Orr H**. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics **144**, 1331–1335.

**Orr H, Presgraves D**. 2000. Speciation by postzygotic isolation: Forces, genes and molecules. BioEssays **22**, 1085–1094.

**Oshlack A, Robinson MD, Young MD**. 2010. From RNA-seq reads to differential expression results. Genome Biology **11**, 220.

**Palmieri N, Schlötterer C**. 2009. Mapping accuracy of short reads from massively parallel sequencing and the implications for quantitative expression profiling. PLoS One **4**, e6323.

**Park HC, Lee S, Park B, *et al.*** 2014. Pathogen associated molecular pattern (PAMP)-triggered immunity is compromised under C-limited growth. Molecules and Cells **38**, 40–50.

**Passardi F, Cosio C, Penel C, Dunand C**. 2005. Peroxidases have more functions than a Swiss army knife. Plant Cell Reports **24**, 255–265.

**Peace C, Bassil N, Main D, *et al.*** 2012. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS One **7**, e48305.

**Peiffer JA, Kaushik S, Sakai H, Arteaga-Vazquez M, Sanchez-Leon N, Ghazal H, Vielle-Calzada JP, Meyers BC**. 2008. A spatial dissection of the *Arabidopsis* floral transcriptome by MPSS. BMC Plant Biology **8**, 43.

**Peng J, Richards DE, Hartley NM, *et al.*** 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature **400**, 256–261.

**Pereira-Lorenzo S, Ramos-Cabrer A, Fischer M**. 2009. Breeding Apple (*Malus* x *domestica* Borkh.). *In*: Jain SM, Priyadarshan PM, (Eds). *Breeding Plantation Tree Crops: Temperate species*. New York: Springer. pp. 33–81.

**Pessina S, Palmieri L, Bianco L, *et al.*** 2017. Frequency of a natural truncated allele of MdMLO19 in the germplasm of *Malus domestica*. Molecular Breeding **37**, 7.

**Petersen R, Krost C**. 2013. Tracing a key player in the regulation of plant architecture: The columnar growth habit of apple trees (*Malus* × *domestica*). Planta **238**, 1–22.

**Petri JL, Hawerroth FJ, Fazio G, Francescatto P, Leite GB**. 2019. Advances in fruit crop propagation in Brazil and worldwide – apple trees. Revista Brasileria de Fruticultura **41**, e-004.

**Di Pierro EA Di, Gianfranceschi L, Guardo M Di, *et al.*** 2016. A high-density, multi-parental

SNP genetic map on apple validates a new mapping approach for outcrossing species. Horticulture Research, 16057.

**Piétu G, Mariage-Samson R, Fayein NA, *et al.*** 1999. The genexpress IMAGE knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. Genome Research **9**, 195–209.

**Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Visser RGF, van de Weg E**. 2014. 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree Genetics and Genomes **10**, 261–271.

**Pilcher RLR, Celton J, Gardiner SE**. 2008. Genetic markers linked to the dwarfing trait of apple rootstock 'Malling 9'. Journal of the American Society for Horticultural Science **133**, 100–106.

**Portin P**. 2014. The birth and development of the DNA theory of inheritance: Sixty years since the discovery of the structure of DNA. Journal of Genetics **93**, 293–302.

**Porto DD, Renou J-P, Bruneau M, Perini P, Anzanello R, Fialho FB, Santos HP dos, Revers LF**. 2015. Transcription profiling of the chilling requirement for bud break in apples: a putative role for FLC-like genes. Journal of Experimental Botany **66**, 2659–2672.

**Potter D, Eriksson T, Evans RC, *et al.*** 2007. Phylogeny and classification of *Rosaceae*. Plant Systematics and Evolution **266**, 5–43.

**Potter D, Gao F, Bortiri PE, Oh SH, Baggett S**. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. Plant Systematics and Evolution **231**, 77–89.

**Pozhitkov AE, Tautz D, Noble PA**. 2007. Oligonucleotide microarrays: Widely applied-poorly understood. Briefings in Functional Genomics and Proteomics **6**, 141–148.

**Rafalski A**. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology **5**, 94–100.

**Rasal KD, Chakrapani V, Pandey AK, Rasal AR, Sundaray JK, Ninawe A, Jayasankar P**. 2017. Status and future perspectives of single nucleotide polymorphisms (SNPs) markers in farmed fishes: Way ahead using next generation sequencing. Gene Reports **6**, 81–86.

**Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z**. 2017. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Molecular Plant **10**, 1047–1064.

**Rédei GP**. 2008. Stutter Bands. *In*: Encyclopedia of genetics, genomics, proteomics and informatics. Dordrecht: Springer Netherlands, 1893.

**Rico C, Cuesta JA, Drake P, Macpherson E, Bernatchez L, Marie AD**. 2017. Null alleles are ubiquitous at microsatellite loci in the Wedge Clam (*Donax trunculus*). PeerJ **5**, e3188.

**Rieseberg LH, Blackman BK**. 2010. Speciation genes in plants. Annals of Botany **106**, 439–455.

**Rieseberg LH, Willis JH**. 2007. Plant speciation. Science **317**, 910–914.

**Risch N**. 1992. Genetic linkage : Interpreting LOD scores. Science **255**, 803–804.

**Robinson JP, Harris SA, Juniper BE**. 2001. Taxonomy of the genus *Malus* Mill. (*Rosaceae*) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Systematics and Evolution **226**, 35–58.

**Robinson MD, McCarthy DJ, Smyth GK**. 2010. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**, 139–140.

**Sachidanandam R, Weissman D, C Schmidt S, M Kakol J, D Stein L, Marth G, Sherry S, C Mullikin J, J Mortimore B, L Willey D**. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**, 928–933.

**Sansavini S, Donati F, Costa F, Tartarini S**. 2004. Advances in apple breeding for enhanced fruit quality and resistance to biotic stresses: New varieties for the European market. Journal of Fruit and Ornamental Plant Research (Special edition) **12**, 13–52.

**Sarowar S, Zhao Y, Soria-Guerra RE, Ali S, Zheng D, Wang D, Korban SS**. 2011. Expression profiles of differentially regulated genes during the early stages of apple flower infection with *Erwinia amylovora*. Journal of Experimental Botany **62**, 4851–4861.

**Sassa H, Hirano H, Ikehashi H**. 1992. Self-incompatibility-related RNases in styles of Japanese Pear (*Pyrus serotina*Rehd .). Plant Cell Physiology **33**, 811–814.

**Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T**. 2007. S locus F-box brothers: Multiple and pollen-specific F-box genes with *S* haplotype-specific polymorphisms in apple and Japanese pear. Genetics **175**, 1869–1881.

**Schaffer RJ, Friel EN, Souleyre EJF, *et al.*** 2007. A genomics approach reveals that aroma production in apple is controlled by ethylene redominantly at the final step in each biosynthetic pathway. Plant Physiology **144**, 1899–1912.

**Schena M, Shalon D, Davis R, Brown P**. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**, 467–470.

**Schippers JHM, Nguyen HM, Lu D, Schmidt R, Mueller-Roeber B**. 2012. ROS homeostasis

during development: An evolutionary conserved strategy. Cellular and Molecular Life Sciences **69**, 3245–3257.

**Schlötterer C**. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma **109**, 365–371.

**Schmidt R, Schippers JHM**. 2015. ROS-mediated redox signaling during cell differentiation in plants. Biochimica et Biophysica Acta - General Subjects **1850**, 1497–1508.

**Segonzac C, Zipfel C**. 2011. Activation of plant pattern-recognition receptors by bacteria. Current Opinion in Microbiology **14**, 54–61.

**Semagn K, Bjoernstad A, Ndjiondjop M**. 2006. Principles, requirements and prospects of genetic mapping in plants. African Journal of Biotechnology **5**, 2569–2587.

**Senan S, Kizhakayil D, Sasikumar B, Sheeja TE**. 2014. Methods for development of microsatellite markers: An overview. Notulae Scientia Biologicae **6**, 1–13.

**Sewelam N, Kazan K, Schenk PM**. 2016. Global plant stress signaling: Reactive oxygen species at the cross-road. Frontiers in Plant Science **7**, 1–21.

**Shiraki T, Kondo S, Katayama S, *et al.*** 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proceedings of the National Academy of Sciences of the United States of America **100**, 15776–15781.

**Shulaev V, Korban SS, Sosinski B, *et al.*** 2008. Multiple models for *Rosaceae* genomics. Plant Physiology **147**, 985–1003.

**Shulaev V, Sargent DJ, Crowhurst RN, *et al.*** 2011. The genome of woodland strawberry (*Fragaria vesca*). Nature Genetics **43**, 109–116.

**Sikuka W**. 2019. South African deciduous fruit exports continue positive growth. USDA Foreign Agricultural Service. Global Agricultural Information Network. Gain Report SA1914.

**Silfverberg-Dilworth E, Matasci CL, Van De Weg WE, *et al.*** 2006. Microsatellite markers spanning the apple (*Malus* x *domestica* Borkh.) genome. Tree Genetics and Genomes **2**, 202–224.

**Soglio V, Costa F, Molthoff JW, Weemen-Hendriks WMJ, Schouten HJ, Gianfranceschi L**. 2009. Transcription analysis of apple fruit development using cDNA microarrays. Tree Genetics and Genomes **5**, 685–698.

**Soneson C, Love MI, Robinson MD**. 2016. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. F1000Research **4**, 1521.

**Soria-Guerra R, Rosales-Mendoza S, Gasic K, Wisniewski M, Band M, S. Korban S**. 2011. Gene expression is highly regulated in early developing fruit of apple. Plant Molecular Biology Reporter **29**, 885–897.

**Stam P**. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant Journal **3**, 739–744.

**Steijger T, Abril JF, Engström PG, *et al.*** 2013. Assessment of transcript reconstruction methods for RNA-seq. Nature Methods **10**, 1177–1184.

**Stephenson AG**. 1981. Flower and fruit abortion: proximate causes and ultimate functions. Annual Review of Ecology and Systematics **12**, 253–279.

**Sun R, Chang Y, Yang F, Wang Y, Li H, Zhao Y, Chen D, Wu T, Zhang X, Han Z**. 2015. A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. BMC Genomics **16**, 1–15.

**Takayama S, Isogai A**. 2005. Self-incomptibility in plants. Annual Review of Plant Biology **56**, 467–489.

**Takos AM, Jaffe FW, Jacob SR, Bogs J, Robinson SP, Walker AR**. 2006. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiology **142**, 1216–1232.

**Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A**. 2011. Differential expression in RNA-seq: a matter of depth. Genome research **21**, 2213–2223.

**Tautz D**. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic acids research **17**, 6463–6471.

**Teneva A, Tomlekova N, Goujgoulova G**. 2014. Major features, mutation mechanism and development of microsatellites as genetic markers. Bulgarian Journal of Agricultural Science **20**, 949–956.

**Thomma BPHJ, Nürnberger T, Joosten MHAJ**. 2011. Of PAMPs and effectors: The blurred PTI-ETI dichotomy. Plant Cell **23**, 4–15.

**Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L**. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology **28**, 511–515.

**Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE**. 2012. Apple, from genome to breeding. Tree Genetics and Genomes **8**, 509–529.

**Troggio M, Šurbanovski N, Bianco L, *et al.*** 2013. Evaluation of SNP data from the *Malus* infinium array identifies challenges for genetic analysis of complex genomes of polyploid origin. PLoS One **8**, e67407.

**Tworkoski T, Miller S**. 2007. Rootstock effect on growth of apple scions with different growth habits. Scientia Horticulturae **111**, 335–343.

**Urrestarazu J, Muranty H, Denancé C, *et al.*** 2017. Genome-wide association mapping of flowering and ripening periods in apple. Frontiers in Plant Science **8**, 1923.

**USDA-FAS**. 2019. Fresh apples, grapes, and pears: World market and trade. United States Department of Agriculture - Foreign Agricultural Service. Washington, DC: USDA. June.

**Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H**. 2003. Structural and transcriptional analysis of the self-incompatibility locus of almond: Identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. The Plant Cell **15**, 771–781.

**Ushijima K, Sassa H, Hirano H**. 1998. Characterization of the flanking regions of the S-RNase genes of Japanese pear (*Pyrus serotina*) and apple (*Malus* x *domestica*). Gene **211**, 159–167.

**Vanderzande S, Micheletti D, Troggio M, Davey MW, Keulemans J**. 2017. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. Tree Genetics and Genomes **13**, 125.

**Varshney R, Graner A, Sorrells M**. 2005. Genic microsatellite markers in plants: Features and applications. Trends in Biotechnology **23**, 48–55.

**Velasco R, Zharkikh A, Affourtit J, *et al.*** 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genetics **42**, 833–839.

**Velculescu VE, Zhang L, Vogelstein B, Kinzler KW**. 1995. Serial analysis of gene expression. Science **270**, 484–487.

**Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Hieter P, Vogelstein B, Kinzler KW**. 1997. Characterization of the yeast transcriptome. Cell **88**, 243–251.

**Verde I, Abbott AG, Scalabrin S, *et al.*** 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nature Genetics **45**, 487–494.

**Verde I, Bassil N, Scalabrin S, *et al.*** 2012. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS One **7**, e35668.

**Vieira MLC, Santini L, Diniz AL, Munhoz C de F**. 2016. Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology **39**, 312–328.

**Viguera E, Canceill D, Ehrlich SD**. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO Journal **20**, 2587–2595.

**Volk GM, Richards CM, Henk AD, Street SM, Collins F, Miller DD, Forsline PL, Genetic P, Unit R**. 2009. Novel diversity identified in a wild apple population from the Kyrgyz Republic. HortScience **44**, 516–518.

**Vu TN, Deng W, Trac QT, Calza S, Hwang W, Pawitan Y**. 2018. A fast detection of fusion genes from paired-end RNA-seq data. BMC genomics **19**, 786.

**Wagner GP, Kin K, Lynch VJ**. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences **131**, 281–285.

**Wang Z, Gerstein M, Snyder M**. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics **10**, 57–63.

**Wang Y, Li W, Xu X, Qiu C, Wu T, Wei Q, Ma F, Han Z**. 2019. Progress of apple rootstock breeding and its use. Horticultural Plant Journal. https://doi.org/10.1016/j.hpj.2019.06.001.

**Wang N, Liu W, Zhang T, Jiang S, Xu H, Wang Y, Zhang Z, Wang C, Chen X**. 2018. Transcriptomic analysis of red-fleshed apples reveals the novel role of MdWRKY11 in flavonoid and anthocyanin biosynthesis. Journal of Agricultural and Food Chemistry **66**, 7076–7086.

**Way R, Lamb R, Pratt C, Cummins J**. 1976. Pale green lethal gene in apple clones. Journal of American Society of Horticultural Science **101**, 676–684.

**Webster AD**. 1995. Rootstock and interstock effects on deciduous fruit tree vigour, precocity, and yield productivity. New Zealand Journal of Crop and Horticultural Science **23**, 373–382.

**Webster T**. 2002. Dwarfing rootstocks: Past, present and future. Compact Fruit Tree **35**, 67-72 (Abstract).

**Wisniewski M, Bassett C, Norelli J, Macarisin D, Artlip T, Gasic K, Korban S**. 2008. Expressed sequence tag analysis of the response of apple (*Malus* x *domestica* 'Royal Gala') to low temperature and water deficit. Physiologia Plantarum **133**, 298–317.

**Wu TD, Nacu S**. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics **26**, 873–881.

**Wu J, Wang Z, Shi Z, *et al.*** 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome

Research **23**, 396–408.

**Wu H, Wang C, Wu Z**. 2012. A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. Biostatistics **14**, 232–243.

**Xie Y, Wu G, Tang J, *et al.*** 2014. SOAPdenovo-Trans: *De novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics **30**, 1660–1666.

**Xu S**. 2013. Genetic mapping and genomic selection using recombination breakpoint data. Genetics **195**, 1103–1115.

**Xue Y, Carpenter R, Dickinson HG, Coen ES**. 1996. Origin of allelic diversity in *Antirrhinum S* locus *RNases*. The Plant Cell **8**, 805–814.

**Yamaguchi S**. 2008. Gibberellin metabolism and its regulation. Journal of Plant Growth Regulation **59**, 225–251.

**Zerbino DR, Birney E**. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Research **18**, 821–829.

**Zhang Q, Ma B, Li H, *et al.*** 2012. Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genomics **13**, 537.

**Zhang Y, Zhu Y, Peng Y, Yan D, Li Q, Wang J, Wang L, He Z**. 2008. Gibberellin homeostasis and plant height control by EUI and a role for gibberellin in root gravity responses in rice. Cell Research **18**, 412–421.

**Zhu Y, Barritt BH**. 2008. Md-ACS1 and Md-ACO1 genotyping of apple (*Malus x domestica* Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genetics and Genomes **4**, 555–562.

**Zhu L, Ni W, Liu S, Cai B, Xing H, Wang S**. 2017. Transcriptomics analysis of apple leaves in response to *Alternaria alternata* apple pathotype infection. Frontiers in Plant Science **8**, 22.

# Chapter 3
# Genetic basis for the mode of inheritance and *S*-linkage underlying crinkled dwarf growth habit trait in apple

This chapter is prepared in accordance to the styles of Journal of Experimental Botany, but it is more elaborate for the purpose of the dissertation.

**Z.T.L. Mbulawa[1,2], A.E. van der Merwe[2], K.M. Soeker[1], and K.R. Tobutt[1]**
*[1]ARC Infruitec-Nietvoorbij, Private Bag X5026, Stellenbosch, 7599, South Africa*
*[2]Department of Genetics, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa*

## 3.1 Abstract

Crinkle dwarfs are economically unfavourable due to their undesirable characteristics for fruit production. A typical seedling phenotype associated with crinkled leaves, poor growth, and in some cases lethality, can be a nuisance in breeding programmes, and potentially regarded as a deleterious trait. Thirteen F1 progenies were raised to investigate the inheritance of the crinkle dwarf trait. The primary cross 'McIntosh' x 'M.9', both parents of normal habit, segregated 9:7 for normal *versus* crinkled dwarf which is attributed to two-gene control for which the two parents are heterozygous (*DdEe*). The crinkled dwarf phenotype occurs when either of the two genes is homozygous recessive *D-ee* or *ddE-*. Of the 13 crosses, seven were fully-compatible and six semi-compatible. Four of the six semi-compatible crosses (sharing an *S*-allele) did not segregate for crinkled dwarf phenotypes. The apparent lack of segregation was initially hypothesized to be due to *S*-linkage. Regardless, involvement of *S*-linkage associated with the crinkled dwarf phenotype was investigated by PCR-based consensus and *S*-allele specific primers of the apple *S*-RNase gene. The findings of the study suggest that crinkle dwarf phenotype may not be *S*-linked, but that it could rather be a case of hybrid incompatibility linked to distortion segregation. Additionally, the study determined a total of eight parental *S*-alleles; $S_1$, $S_2$, $S_3$, $S_5$, $S_7$, $S_9$, $S_{10}$, $S_{24}$ and $S_{37}$. The *S*-genotypes of the rootstock cultivar Malling 1'M.1' ($S_3S_9$) and the selection TSR1T187 ($S_7S_{24}$), were deduced

in this study. Moreover, the *S*-genotypes for 'Irish Peach' ($S_1S_{37}$) and 'Howgate Wonder' ($S_3S_5$) have been previously reported in a pilot study conducted at East Malling Research (Bošković, unpublished), and have been confirmed in this study, though $S_5$ and newly detected $S_x$ alleles in 'Howgate Wonder' remains to be resolved. This study has provided insight underlying the genetic basis and the inheritance behind the crinkle dwarf growth habit in apple.

## 3.2 Introduction

Dwarf growth habits are common throughout the plant kingdom (Garvey, 1985; Garvey and Lyrene, 1987; Milach and Federizzi, 2001; Komorisono, 2005; Tworkoski and Miller, 2007; Fazio *et al.*, 2014). Garvey (1985) described two types of dwarf growth habits in plants: complete-dwarfs with all organs smaller than normal and semi-dwarf with shorter internodes. Dwarf types are found, and may be successfully exploited, in many crop plants and the genetic pathways to dwarfism appear to be numerous and diverse (Seleznyova *et al.*, 2008; Bai *et al.*, 2012; Tong *et al.*, 2012). Sun and Gubler (2004) reported that dwarfism in plants is often caused by mutations in genes controlling the biosynthesis or signaling pathways of the plant hormone gibberellin (GA) (Sun and Gubler, 2004). Dwarf growth habit may also arise due to semi-lethal genes or incompatibilities within the genome (Garvey, 1985). Dwarf phenotypes could also be due to aneuploidy because these plants are derived from weak and sub-viable gametes since they are carrying an odd number of chromosomes (Garvey and Lyrene, 1987). Aberrant cellular division or elongation has been reported to cause dwarfism in some mutants such as the rice (Sato *et al.*, 1999; Komorisono, 2005; Yang *et al.*, 2011).

Alston (1976) discovered three rare forms of dwarf types in apple (early, crinkled and sturdy dwarfs). One, crinkle dwarf growth habit, was reported in a cross of 'Irish Peach' x TSR1T187, and the crinkled phenotype attributed to control possibly by a single recessive allele, where both parents are heterozygous. Crinkled dwarf phenotypes are characterised by dwarf seedlings associated with small, rounded, dark-green crinkled leaves, recognisable at four to six weeks after germination but more distinct from twelve weeks (Alston, 1976).

The domesticated apple (*Malus pumila* Mill.; $2n=2x=34$) belongs to the family *Rosaceae* under the subtribe *Pyrinae* of the subfamily *Spiraeoideae* (formerly, *Maloideae*), (Potter *et al.*, 2007).

Apple, like other fruit species of the *Rosaceae*, exhibits a gametophytic self-incompatibility system (GSI), also shared by two other distantly related families, the *Solanaceae* and *Scrophulariaceae* (DeNettancourt, 2001; McClure and Franklin-Tong, 2006). The GSI is one of the most common reproductive systems in flowering plants, and it prevents inbreeding thus promoting outcrossing (DeNettancourt, 2001). The GSI mechanism is genetically attributed to a single, highly polymorphic genetic locus, designated as the *S*-locus (Newbigin *et al.*, 1993; Kao and Tsukamoto, 2004; McClure and Franklin-Tong, 2006; McClure, 2009). The underlying GSI system is based on the interaction of at least two tightly linked gene products, encoding for pistil and pollen, both expressed in a haplotype-specific manner (Sassa *et al.*, 1996; DeNettancourt, 1997; Cheng *et al.*, 2006).

The pistil (female) expressed determinant is a stylar glycoprotein with ribonuclease activity, *S*-RNase, and was initially characterised in *Nicotiana alata* (Anderson *et al.*, 1989; McClure *et al.*, 1990; Newbigin and Uyenoyama, 2005). Since then, the *S*-RNase gene in *Rosaceae* species: e.g. apple (*Malus x domestica*), Japanese pear (*Pyrus pyrifolia*) (Sassa *et al.*, 2007), sweet cherry (*Prunus avium*) (Bošković and Tobutt, 1996; Bošković *et al.*, 1997) and almond *(Prunus dulcis*) (Ushijima *et al.*, 2001; Ortega *et al.*, 2006), amongst others has been fully characterised (Gu *et al.*, 2015; Herrera *et al.*, 2018). The *S*-RNase gene typically consists of a signal peptide region (SP) located at the amino terminus, five conserved regions; C1, C2, C3, RC4 (*Rosaceae*-specific) and C5, which determine the *S*-protein structure and function, and one hypervariable region (HV), which plays an important role in the discrimination of self from nonself pollen (Broothaerts *et al.*, 1995; Sassa *et al.*, 1996; Ishimizu *et al.*, 1998).

The pollen determinant (male) has been identified and consists of multiple pollen-specific *F*-box genes or *S*-locus *F*-box brothers (*SFBB*s), in apple (*Malus* × *domestica*) and Japanese pear (*Pyrus pyrifolia*). Two distinct *SFBB*s genes were identified and characterised in the genomic regions surrounding each of the apple *S*-haplotypes *S*-RNase $S_3$ and $S_9$ (MdSFBBα and MdSFBBβ) from the cultivar 'Florina' and six distinct F-box genes were associated with each of the *S*-haplotypes $S_4$ and $S_5$ ($PpSFBB^{4\text{-}\alpha}$, $PpSFBB^{4\text{-}\beta}$, $PpSFBB^{4\text{-}\gamma}$, $PpSFBB^{5\text{-}\alpha}$, $PpSFBB^{5\text{-}\beta}$, $PpSFBB^{5\text{-}\gamma}$) in Japanese pear, cultivar 'Kosui' (Cheng *et al.*, 2006; Kakui *et al.*, 2007; Sassa *et al.*, 2007; Minamikawa *et al.*, 2010; Okada *et al.*, 2011; Claessen *et al.*, 2019). Recently, additional *SFBB* genes in apple have been identified (Pratas *et al.*, 2018).

Genetic investigation of self-incompatibility has been carried out extensively on many apple cultivars, including traditional pollination studies (Kobel, 1939; Certal *et al.*, 1999), *S*-RNase protein identification and the current cDNA cloning and nucleotide sequencing (Broothaerts *et al.*, 1995; Janssens *et al.*, 1995; Sassa *et al.*, 1996; Verdoodt *et al.*, 1998; Matsumoto *et al.*, 2010; Gu *et al.*, 2015; De Franceschi *et al.*, 2016). Additionally, the *S*-allele-specific PCR-based primers were first developed by Janssens *et al*. (1995). Since then, *S*-allele specific primers have been developed for other related species of the *Rosaceae* (Wu *et al.*, 2013) and currently utilised in almond (Ortega *et al.*, 2006; Martí *et al.*, 2011; Hafizi *et al.*, 2013), sweet cherry (Tao *et al.*, 1999; Sonneveld *et al.*, 2001; Wünsch and Hormaza, 2004), pear (Tao *et al.*, 1997; Ishimizu *et al.*, 1999; Zuccherelli *et al.*, 2002; Minamikawa *et al.*, 2010; Kakui *et al.*, 2011) and apricot (Tao *et al.*, 2000; House, 2007).

The success of the GSI mechanism is dependent on the combination of the *S*-haplotypes (Hiratsuka and Zhang, 2002; Claessen *et al.*, 2019). The use of PCR-based consensus and allele-specific primers are widely utilised to speed up the determination of the *S*-genotypes and facilitating the design of successful crosses with maximum crop set. The crosses are fully-compatible when both the maternal and paternal *S*-genotypic constitution are different e.g. $S_1S_2$ x $S_3S_4$. On the other hand, a semi-incompatible cross arises when the two parents share a common *S*-allele e.g. $S_1S_2$ x $S_1S_3$, and its success is highly governed by the direction of the cross. The degree of self-incompatibility varies in triploid cultivars, resulting in fruit set when pollinated with diploids, but fluctuating considerably when crossed with other triploids. Bošković and Tobutt, (1999) reported that some apple cultivars are incompatible when crossed, and some triploid x diploid combinations fail whereas the reciprocals succeed.

Moreover, differential transmission of paternal gametes in semi-compatible crosses can cause distorted segregation ratios for genes linked to the *S*-locus (Zamir and Tadmor, 1986; O'Leary and Boyle, 1998; Harbord *et al.*, 2000; Rabbani *et al.*, 2012). Dai *et al*. (2017) defined segregation distortion as a deviation of the observed allelic frequencies at a locus from the expected Mendelian ratio in a segregating population.

Currently, little is known of the genetics underlying the crinkled dwarf growth habit phenomena. This study aims to provide a better understanding on the mode of inheritance underlying crinkled dwarf trait. Therefore, thirteen apple F1 progenies were raised from the crosses: 'McIntosh' x

Malling 9 (‘M.9’), ‘McIntosh’ x Malling 1 (‘M.1’), ‘M.1’ x ‘M.9’, ‘Telamon’ x ‘M.1’, ‘Telamon’ x ‘M.9’, ‘McIntosh’ x ‘Telamon’, ‘McIntosh’ x ‘Tuscan, ‘Trajan’ x ‘M.9’, ‘Tuscan’ x ‘M.9’, ‘McIntosh Wijcik’ x ‘Irish Peach’, ‘McIntosh Wijcik’ x TSR1T187, ‘Irish Peach’ x ‘M.9’ and TSR1T187 x ‘M.9’.

The main objectives of this study was first: to investigate the genetic basis of the dwarf growth habit associated with crinkled leaves by studying the phenotypic segregation patterns in the thirteen apple F1 progenies; and sec to investigate the possibilities of *S*-linkage by studying the parental *S*-genotypic constitution using the PCR-based consensus and allele-specific primers of the apple *S*-RNAse gene.

## 3.3 Materials and Methods

Some sections of the materials and methods are similar across the three experimental chapters of this thesis but have been included in each chapter for completeness.

### 3.3.1 Plant materials and establishment of segregating progenies

Thirteen crosses were made with the aim to raise progenies segregating for dwarf seedling with crinkled leaves: ‘McIntosh’ x ‘M.9’, ‘McIntosh’ x ‘M.1’, ‘M.1’ x ‘M.9’, ‘Telamon’ x ‘M.1’, ‘Telamon’ x ‘M.9’, ‘Telamon’ x ‘McIntosh’, ‘Trajan’ x ‘M.9’, ‘Tuscan’ x ‘M.9’, ‘McIntosh’ x ’Tuscan’, ‘McIntosh Wijcik’ x ‘’Irish Peach’, ‘McIntosh Wijcik’ x TSR1T187, ‘Irish Peach’ x ‘M.9’ and TSR1T187 x ‘M.9’. All the crosses pertaining to parentals ‘McIntosh Wijcik’, ’Irish Peach’ and TSR1T187 were kindly generated and provided by K. Tobutt. The details of the crosses, with their respective hypothesized parental genotypes and progeny segregations are detailed in Table 3.1.

The parental cultivars selected in the generation of the crosses in this study are known to carry dwarf genes (K. Tobutt pers. communication); ‘McIntosh’ and its derivates ‘Telamon’, ‘Trajan’, ‘Tuscan’, and dwarfing rootstock cultivars: ‘M.1’, ‘M.9’, and two additional parental cultivars, ‘Irish Peach’, TSR1T187, adapted from Alston (1976), are reported to segregate for the crinkled dwarf trait.

### 3.3.2 Mapping population and growth conditions

All the F1 apple progenies used in this study were raised at Bien Donné Research Farm of the Agricultural Research Council (ARC), Groot Drakenstein, Western Cape, South Africa [(33°83'33"32.06 (S); 18°98'33"33.59 (E)]. Briefly, controlled pollinations of all crosses were conducted in spring of 2015 to 2017, thereafter fruits were harvested and the seeds extracted, washed with distilled water and dried at room temperature. Subsequently, seeds were placed in sealed plastic bags containing moist peat and vermiculite (50:50 w/w) substrate and stored between 0-4°C for a period of 12 weeks, to break dormancy under artificial winter conditions. Thereafter, germinated seeds were sown in trays under glasshouse conditions (~20-24°C) until the seedlings had grown to two-to-four leaf stage. The seedlings were transplated into compost-containing pots and allowed to grow on their own roots in the glasshouse.

### 3.3.3 Phenotypic assessment of crinkled dwarf

All progenies were scored visually for the presence or absence of dwarf seedlings with crinkled leaves, from four weeks after germination and continuing throughout the growing season. The crinkled dwarf phenotype is distinct 12 weeks after germination, where seedlings were scored with confidence (Figure 3.1). The seedlings occasionally show weak growth and develop poorly and may be associated with some lethality, self-incompatibility and other modifier genes. Other morphological characters scored included seedling height and diameter at different heights in the progeny of 'McIntosh' x 'M.1' (Supplementary Table 3.1). The internode length differences were also noted. The segregation ratios for each progeny were computed and compared against Mendelian ratios of 3:1 and/or 9:7 corresponding to their respective hypothesized parental genotypes. Significant deviations were determined using Chi-square test ($\chi 2$) at $\rho<0.05$ conducted in Microsoft Excel (Table 3.1).

**Figure 3.1. Distinct phenotypic segregation of normal *versus* dwarf seedlings with dark-green crinkled leaves 12 weeks after germination**.

### 3.3.4 Plant material, DNA extraction and quantification

Approximately, 300 mg young leaves were collected from all the F1 progeny seedlings and their parents, placed in 2 mL centrifuge tubes and stored at -80°C until required for genomic DNA extraction.

Total genomic DNA was extracted according to the cetyltrimethyl-ammonium-bromide (CTAB) method (Doyle and Doyle, 1987), with slight modification. A single, 3 mm stainless-steel ball bearing was placed in each 2 mL centrifuge tube containing frozen leaf material. Briefly, 800 µL (preheated to 65°C) extraction buffer [2% (w/v) CTAB, 2% (w/v) PVP 40, 1.4 M NaCl, 20 mM EDTA (pH 8), 100 mM Tris-HCl (pH 8)] and 1% (0.8 µL) β-mercaptoethanol were added, and initially mixed by inversion. Subsequently, the centrifuge tubes were homogenised using a Tissuelyser II ball mill (Qiagen, Hilden, Germany) at a frequency of 30 Hz for 2-4 min or until all leaf tissues were completely ground. The sample tubes were incubated at 65°C for 2 hr, with inversion every 15 min. Thereafter, the ball bearings were removed with a stainless steel magnet. Subsequently, 800 µL of chloroform-isoamyl alcohol (24:1 v/v) was added, mixed by inversion and centrifuged at 13500 rpm for 15 min. The top aqueous phase was aliquoted into a new 1.5 mL centrifuge tube, and 600 µL of chloroform-isoamyl alcohol (24:1 v/v) added and the samples centrifuged at 13500 rpm for 10 min. The top aqueous phase was transferred into a new 1.5 mL Eppendorf tube and precipitated with 500 µL cold isopropanol, overnight, at -20°C. The precipitate

was centrifuged at 13500 rpm at 4°C for 15 min. The pellet was washed in 500 µL of ice cold 70% (v/v) ethanol by centrifugation at 13500 rpm at 4°C for 15 min. The pellet was dried and resuspended in 50 µL of 1 x TE buffer (10 mM of Tris-HCl, 0.1 mM of EDTA, pH 8.0).

The quality and quantity of the DNA was determined with a BioDrop spectrophotometer (BioDrop Technology, Rockland, UK), following the manufacturer' instructions. The DNA was diluted to two working concentrations; 10 ng/ µL and 30-40 ng/ µL and subsequently stored at -20°C until further use.

### 3.3.5 *S*-genotyping of parents

The parental *S*-genotypic constitutions were determined by PCR using a consensus primer pair (hereafter referred to as SRB) (R. Bošković, unpublished) and eight allele-specific primers ($S_1$, $S_2$, $S_3$, $S_5$, $S_7$, $S_9$, $S_{10}$ and $S_{24}$) (Long *et al*., 2010). The forward primers were fluorescently labelled with 6-FAM and PET (Applied Biosystems). All the PCR amplifications undertaken in this study were carried out in an Applied Biosystems GeneAmp PCR System 9700 thermal cycler. The nucleotide sequences of the consensus primer set and allele-specific primers used are listed in the Supplementary Table 3.2.

***S-genotyping with PCR-based consensus primers***

The PCR conditions were optimised from the protocol designed to detect S-alleles in apple (R. Bošković, unpublished). Briefly, PCR conditions: RNase-free $H_2O$ (7.9 µL), 25mM $MgCl_2$ (3.0 µL), Flexi Taq green buffer (4.0 µL), 10mM dNTPs (0.5 µL), 5U GoTaq polymerase (0.1 µL) (Promega, Madison, Wisconsin, USA), 30-40 ng DNA (2.5 µL), 10mM SRB (F) (1 µL), 10mM SRB (R) (1 µL). PCR reactions were performed using the following cycling profile: 3 min of denaturation at 94°C; 10 cycles of 10 sec at 94°C, 2 min at 55°C, 2 min at 72°C, followed by 25 cycles (with 10 sec increment per cycle to the extension step) of 10 sec at 94°C, 2 min at 55°C, 2 min at 72°C, and a final extension of 10 min at 68°C, followed by cooling to 4°C.

**S-*genotyping with PCR-based allele-specific primers***

The parental *S*-genotypes were further resolved with the allele-specific primers. The PCR amplification was conducted in a final volume of 12.5 µL containing 30-40 ng DNA, 6.25 µL of

Qiagen multiplex kit, 1 µL (10 µM) of each forward and reverse allele-specific primer and 2.75 µL of RNase-free water. The PCR conditions were: an initial denaturation at 95°C for 15 min, followed by 29 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 1 min, with a final 15 min extension at 72°C, followed by cooling to 4°C.

***Agarose gel eletrophoresis***

The PCR products amplified with both the consensus primers and allele-specific primers, were resolved on a 2 % (w/v) agarose gel containing ethidium bromide and eletrophoresed at 50 V for 5.5 hr in a 1x TBE buffer (0.089 M Tris Base, 0.089 M Boric acid, and 0.002 M EDTA). The fragments were visualised on a UV transilluminator, photographed and digitised using the Syngene Ingenius LHR gel documentation system with GeneSnap™ version 7.12.06 (Syngene, Cambridge, United Kingdom). Fragment lengths of the amplified fragments were estimated using GeneRuler 1 kb plus-DNA ladder (Thermo Scientific). The PCR amplicons were also sized on an automated ABI PRISM 3500 capillary analyser (Applied Biosystems) at the Stellenbosch University Central Analytical Facilities (CAF) DNA Sequencer Unit. The fragment analysis visualisations were established in relation to the internal size standard, GS1200LIZ in GeneMapper version 5.0 software (Applied Biosystems). The allelic sizes were validated independently by a competent co-worker and collated in spreadsheets for further reference.

## 3.4 Results and Discussion

### 3.4.1 Morphological characteristics

The leaf morphology was visually scored as either “normal”, healthy leaves, or “crinkle”, abnormal dark green crinkled leaves. The morphological characteristic based on plant height and diameter were measured at 36 weeks after germination with the aim to distinguish between normal *vs* crinkle dwarf growth characteristics. The plant height was measured from the ground to the edge of the tip. On average, normal seedlings were taller at 57.41 cm and crinkle dwarf phenotypes measured at 11.24 cm. The plant height between normal *vs* crinkle dwarfs was roughly a difference of 5 fold. The diameter of the seedlings at breast height and diameter at middle point showed a clear distinction, which were of 3 and 2 fold differences respectively. In contract, the diameter measure at tip were relatively comparable between normal and crinkle dwarf (Supplementary Table 3.1).

### 3.4.2 Phenotypic segregation analysis of F1 progenies

The rationale and test of significance ($\chi 2$ at $\rho<0.05$) on the inheritance of crinkled dwarf phenotype through segregation analysis of each progeny are listed in Table 3.1 and the *priori* *S*-linkage hypotheses on semi-compatible crosses are presented in Table 3.2.

**Table 3.1. Segregation of 13 apple F1 progenies used in this study. Hypothetical maternal and paternal genotypes, with respective observed and expected crinkled dwarf segregation ratios, and chi-square $\chi^2$ ($\rho<0.05$) significance indicated for each F1 progeny.**

| Hypothetical parental genotypes | | Total no. seedlings | Crinkled dwarf segregation | | $\chi$2 ($\rho<0.05$) | |
|---|---|---|---|---|---|---|
| **Female parent ♀** | **Male parent ♂** | | **Observed** | **Expected** | **3:1** | **9:7** |
| **McIntosh** (Ho: *DdEe*) | **M.9** (Ho: *DdEe*) | 150 | 83 : 67 | ~9 : 7 | n.c. | 0.051 (0.82) |
| **McIntosh** (Ho: *DdEe*) | **M.1** (Ho: *DDEe*/*DdEE*) | 118 | 90 : 28 | ~3 : 1 | 0.10 (0.750) | 19.22 ($1.16x10^{-5}$) |
| **M.1** (Ho: *DDEe*/*DdEE*) | **M.9** (Ho: *DdEe*) | 44 | 44 : 0 | ~3 : 1 | n.c. | n.c. |
| **Telamon** (Ho: *DDEe*/*DdEE* or *DdEe* ) | **M.1** (Ho: *DDEe*/*DdEE*) | 79 | 62 : 12 | ~3 : 1 or 9 : 7 | 3.05 (0.080) | 22.79 ($1.8x10^{-6}$) |
| **Telamon** (Ho: *DDEe*/*DdEE* or *DdEe* ) | **M.9** (Ho: *DdEe*) | 24 | 14 : 10 | ~3 : 1 or 9 : 7 | 3.56 (0.059) | 0.04 (0.837) |
| **Trajan** (Ho: *DDEe*/*DdEE* or *DdEe*) | **M.9** (Ho: *DdEe*) | 15 | 6 : 9 | ~3 : 1 or 9 : 7 | 9.80 (0.002) | 1.61 (0.204) |
| **Tuscan** (Ho: *DDEe*/*DdEE* or *DdEe*) | **M.9** (Ho: *DdEe*) | 21 | 13 : 8 | ~3 : 1 or 9 : 7 | 1.92 (0.166) | 0.27 (0.601) |
| **McIntosh** (Ho: *DdEe*) | **Telamon** (Ho: *DDEe*/*DdEE* or *DdEe* ) | 23 | 23 : 0 | ~3 : 1 or 9 : 7 | n.c. | n.c. |
| **McIntosh** (Ho: *DdEe*) | **Tuscan** (Ho: *DDEe*/*DdEE* or *DdEe*) | 90 | 90 : 0 | ~3 : 1 or 9 : 7 | n.c. | n.c. |
| **McIntosh Wijcik** (Ho: *DdEe*) | **Irish Peach** (Ho: *DDEe*/*DdEE* or *DdEe*) | 113<br>44 | 79 : 27<br>34 : 10 | ~3 : 1 or 9 : 7 | 0.01 (0.911)<br>0.12 (0.728) | 14.39 (0.00)<br>7.90 (0.004) |
| **McIntosh Wijcik** (Ho: *DdEe*) | **TSR1T187** (Ho: *DDEe*/*DdEE* or *DdEe*) | 7 | 48 : 11 | ~3 : 1 or 9 : 7 | 1.27 (0.260) | 15.11(0.001) |
| **Irish Peach** (Ho: *DDEe*/*DdEE* or *DdEe*) | **M.9** (Ho: *DdEe*) | 25 | 19 : 0 | ~3 : 1 or 9 : 7 | n.c. | n.c. |
| **TSR1T187** (Ho: *DDEe*/*DdEE* or *DdEe*) | **M.9** (Ho: *DdEe*) | 43<br>24 | 24 : 19<br>18 : 5 | ~3 : 1 or 9 : 7 | 8.44 (0.004) | 0.003 (0.954) |

n.c. =not calculated

**Table 3.2. Parental S-genotypes and expected progeny S-genotypic segregation classes as per priori S-linkage hypothesis.**

| **Female parent ♀** | **Male parent ♂** | **Parental *S*-genotypes** | **Expected progeny *S*-genotypic classes** | **Cross-compatibility** |
|---|---|---|---|---|
| McIntosh | M.9 | $S_{10}S_{25}$ x $S_1S_3$ | $S_1S_{10}$ : $S_1S_{25}$ : $S_3S_{10}$ : $S_3S_{25}$ | fully-compatible |
| McIntosh | M.1 | $S_{10}S_{25}$ x $^1\boldsymbol{S_3S_9}$ | $S_3S_{10}$ : $S_3S_{25}$ : $S_9S_{10}$ : $S_9S_{25}$ | fully-compatible |
| M.1 | M.9 | $^1\boldsymbol{S_3S_9}$ x $S_1S_3$ | $S_1S_3$ : $S_1S_9$ | semi-compatible |
| Telamon | M.1 | $S_3S_{10}$ x $^1S_3S_9$ | $S_3S_9$ : $S_9S_{10}$ | semi-compatible |
| Telamon | M.9 | $S_3S_{10}$ x $S_1S_3$ | $S_1S_3$ : $S_1S_{10}$ | semi-compatible |
| McIntosh | Telamon | $S_{10}S_{25}$ x $S_3S_{10}$ | $S_3S_{25}$ : $S_3S_{10}$ | semi-compatible |
| McIntosh | Tuscan | $S_{10}S_{25}$ x $S_5S_{10}$ | $S_5S_{25}$ : $S_5S_{10}$ | semi-compatible |
| Trajan | M.9 | $S_2S_{25}$ x $S_1S_3$ | $S_1S_2$ : $S_1S_{25}$ : $S_2S_3$ : $S_3S_{25}$ | fully-compatible |
| Tuscan | M.9 | $S_5S_{10}$ x $S_1S_3$ | $S_1S_5$ : $S_1S_{10}$ : $S_3S_5$ : $S_3S_{10}$ | fully-compatible |
| McIntosh Wijcik | Irish Peach | $S_{10}S_{25}$ x $^2\boldsymbol{S_1S_{37}}$ | $S_1S_{10}$ : $S_{10}S_{37}$ : $S_1S_{25}$ : $S_{25}S_{37}$ | fully-compatible |
| McIntosh Wijcik | TSR1T187 | $S_{10}S_{25}$ x $^1\boldsymbol{S_7S_{24}}$ | $S_7S_{10}$ : $S_{10}S_{24}$ : $S_7S_{25}$ : $S_{24}S_{25}$ | fully-compatible |
| Irish Peach | M.9 | $^2\boldsymbol{S_1S_{37}}$ x $S_1S_3$ | $S_1S_3$ : $S_3S_{37}$ | semi-compatible |
| TSR1T187[1] | M.9 | $^1\boldsymbol{S_7S_{24}}$ x $S_1S_3$ | $S_1S_7$ : $S_3S_7$ : $S_1S_{24}$ : $S_3S_{24}$ | fully-compatible |

[1]S-genotype for 'M.1' ($\boldsymbol{S_3S_9}$) and TSR1T187 ($\boldsymbol{S_7S_{24}}$) deduced in this study; [2]S-genotype for 'Irish Peach (($\boldsymbol{S_1S_{37}}$)' together with a reference 'Howgate Wonder'($\boldsymbol{S_3S_5S_x}$) were confirmed in this study

### 3.4.3 S-genotypic determination of parents

The observed distorted segregation of normal *versus* crinkled dwarf phenotypes in some of the progenies (Table 3.1), indicated a possibility of the crinkled dwarf trait being *S*-linked. The *S*-linkage became apparent when the crosses: 'M.1' ($S_3S_9$) x 'M.9' ($S_1S_3$), 'McIntosh' ($S_{10}S_{25}$) x 'Telamon $S_3S_{10}$)', 'McIntosh' ($S_{10}S_{25}$) x 'Tuscan' ($S_5S_{10}$) and 'Irish Peach' ($S_1S_{37}$) x 'M.9' ($S_1S_3$), with parents sharing a common *S*-allele i.e. semi-compatible crosses, did not segregate for crinkled dwarfs phenotype. Contrary to this, the other two semi-compatible crosses; 'Telamon' ($S_3S_{10}$) x 'M.1' ($S_3S_9$) and 'Telamon' ($S_3S_{10}$) x 'M.9' ($S_1S_3$), segregated for crinkled dwarfs.

Prior to raising the progenies, the S-genotype of the parents and the direction of the parental cross were not taken into consideration. Fortunately, the *S*-genotype status for the parentals: 'McIntosh', 'McIntosh Wijcik' (Van Nerum *et al.*, 2001; Kitahara and Matsumoto, 2002*a*,*b*), 'Telamon', 'Tuscan', 'Trajan'(Broothaerts *et al.*, 2004), and 'M.9'(Agapito-Tenfen *et al.*, 2015) are publicly available, and that of 'Irish Peach', 'Howgate Wonder' (Bošković, unpublished). The *S*-genotypes for 'M.1' and 'TSR1T187' were deduced in the current study. The S-genotypes of all the parental cultivars used in this study were successfully determined using the consensus and allele-specific primers of the apple S-RNase gene (Bošković, unpublished; Long *et al.*, 2010). The apple cultivars: 'Gala' ($S_2$, $S_5$) (Janssens *et al.*, 1995; Dreesen *et al.*, 2010) and 'Golden Delicious' ($S_2$, $S_3$) (Broothaerts *et al.*, 2004; Dreesen *et al.*, 2010; Long *et al.*, 2010) were used as controls.

The sizes of the PCR amplification products were initially determined on agarose gel against a 1Kb plus gene ruler marker (Figures 3.2 and 3.3). To confirm allele sizes, fragment analysis was performed on an ABI capillary sequencer using fluorescently labelled consensus and allele-specific primers (Table 3.3).

#### S-*genotyping with consensus primer*

Initially, consensus primers of the apple *S*-RNase gene designed by Bošković (unpublished) (Supplementary Table 3.2) was used to amplify the *S*-alleles from the genomic DNA of

'McIntosh' ($S_{10}S_{25}$) (Broothaerts *et al.*, 2004; Dreesen *et al.*, 2010), 'Telamon' ($S_3S_{10}$), Tuscan ($S_5S_{10}$), Trajan ($S_2S_{25}$) (Broothaerts *et al.*, 2004), 'M.1' ($S_3S_9$) (this study) , 'M.9' ($S_1S_3$), 'Irish Peach' ($S_1S_{37}$) (R. Bošković, unpublished), TSR1T187 ($S_7S_{24}$) (this study). The four apple cultivars: 'Adams Pearmain' ($S_1S_3S_{10}$), 'Gala' ($S_2S_5$), 'Golden Delicious' ($S_2S_3$) (Broothaerts *et al.*, 2004; Dreesen *et al.*, 2010; Long *et al.*, 2010) and 'Howgate Wonder' ($S_3S_5$) (R. Bošković, unpublished) were used as controls for PCR.

Most of the *S*-alleles were successfully amplified with the consensus primer, and were in accordance with the *S*-allele sizes reported in R. Bošković (unpublished). Allele sizes ranged from 230 bp to 1500 bp (Table 3.3). The consensus primers were also successful in identifying the *S*-alleles of the cultivars and selection not previously reported i.e. $S_7$, $S_9$, $S_{24}$ $S_{37}$ and $S_x$ alleles. Initially, the $S_9$ (258 bp) and $S_{24}$ (448 bp) bands were scored incorrectly as they could not be discriminated from $S_2$ (262 bp) and $S_1$ (454 bp) with similar sizes, respectively. There were no detectable *S*-alleles with sizes above 1500 bp i.e. $S_{10}$ could not be identified using the consensus primers. This was expected for cultivars 'McIntosh', 'Telamon' and 'Tuscan' with fragment sizes of 1800 bp and for $S_{25}$ at 2600 bp for 'McIntosh' and 'Trajan'. Additionally, a PCR amplicon of 'Howgate Wonder' ($S_3S_5$), resulted in diffused bands (Figure 3.2) representing $S_3S_5$ i.e. 1400 bp and 1300 bp, respectively, with a third distinct band of 933 bp, assigned $S_x$. The PCR amplicons obtained with consensus primers were then further analysed on a capillary sequencer in determining the exact fragment sizes of the bands obtained in agarose gel (Table 3.3).

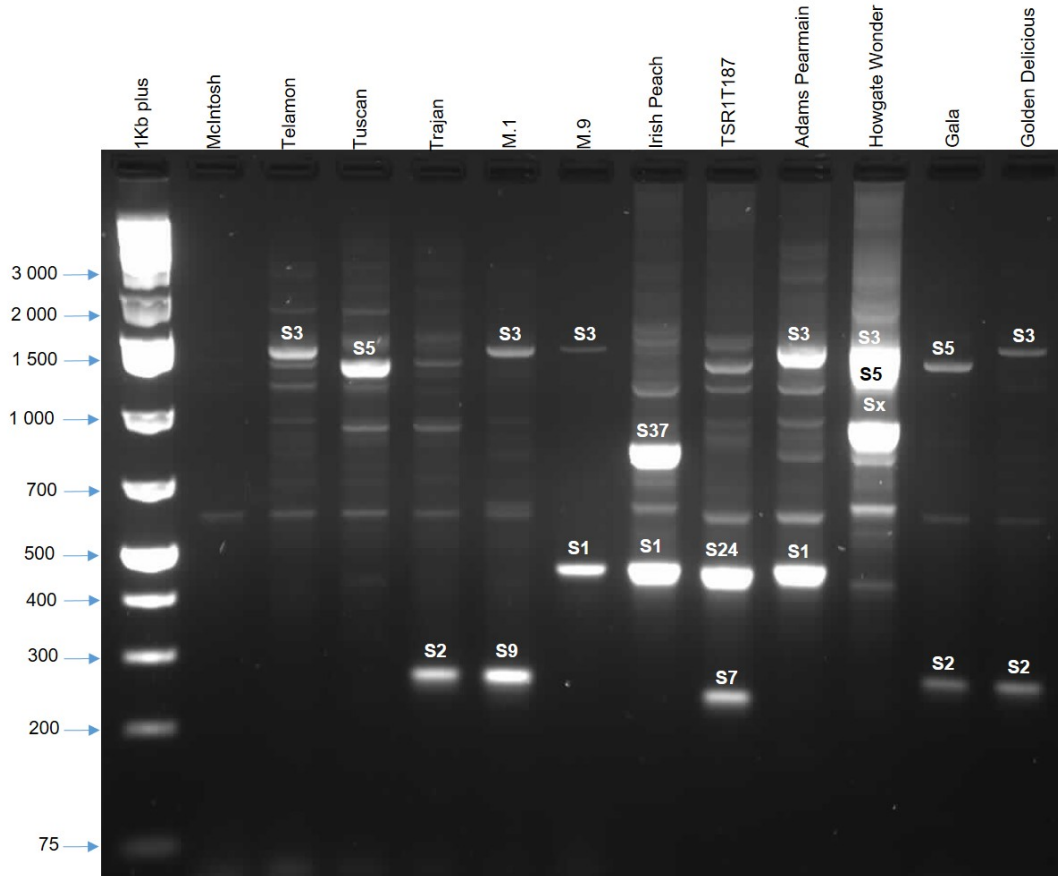


**Figure 3.2. Electrophoretic separation on a 2% agarose gel showing amplified parental S-allele band patterns with consensus primer set of the apple S-RNase gene.** Each lane shows the outcome of a single allele-specific PCR amplicon as indicated. 1Kb plus denotes a GeneRuler 1 kb plus-DNA ladder used as a molecular size marker estimator in base pairs.

### S-*genotyping with allele-specific primers*

The allele-specific primers resulted in distinct bands on agarose gel, and were in accord with the allele sizes reported in Long *et al*., 2010 (Table 3.3). However, amplification was not successful for some *S*-alleles i.e. $S_{1?}$, $S_{2?}$, $S_{10?}$ for TSR1T187, 'M.1'and 'Adams Pearmain', as illustrated in Figure 3.3, which indicated the absence of the *S*-allele. These results have validated the correct scoring of the $S_7$, $S_9$ and $S_{24}$ alleles which were initially scored incorrectly with the consensus primer set. These results led to the *S*-genotypic determination of TSR1T187 and 'M.1'to be $S_7S_{24}$, and $S_3S_9$, respectively. To our knowledge this is the first study to report the *S*-genotypes of the apple selection TSR1T187 and the rootstock cultivar,

'M.1'.The *S*-genotypes of 'Irish Peach' ($S_1S_{37}$) and 'Howgate Wonder' ($S_3S_5$). In literature, 'Adam Pearmain' is reported with three *S*-genotypes; $S_1S_3S_{10}$ (Broothaerts *et al.*, 2004; Dreesen *et al.*, 2010). In this study $S_{10}$ (denoted as $S_{10?}$ in Figure 3.3) could not be amplified. The $S_{25}$ and $S_{37}$ bands could not be verified with the allele-specific primers, as the primers remains to be designed.

**Table 3.3. S-allele identified using consensus and allele-specific primers.**

| ***S*-alleles** | **Consensus (bp)** | | **Allele-specific (bp)** | |
|---|---|---|---|---|
| | **[1]Bošković (bp)** | **This study (bp)** | **Long *et al.*, 2010** | **This study (bp)** |
| $S_1$ | 450 | 454 | 734 | 705 |
| $S_2$ | 259 | 262 | 489 | 494 |
| $S_3$ | 1400 | **~1400** | 292 | 287 |
| $S_5$ | 1300 | **~1300** | 1447 | **~1447** |
| $S_7$ | 231 | 233 | 397 | 391 |
| $S_9$ | 256 | 258 | 522 | 517 |
| $S_{10}$ | 1800 | n.a | 203 | 198 |
| $S_{24}$ | 445 | 448 | 421 | 401/416 |
| $S_{25}$ | 2600 | n.a | - | - |
| $S_{37}$ | n.p.k | 828 | - | - |

[1] Bošković (unpublished); n.a = not amplified; n.p.k = no prior knowledge

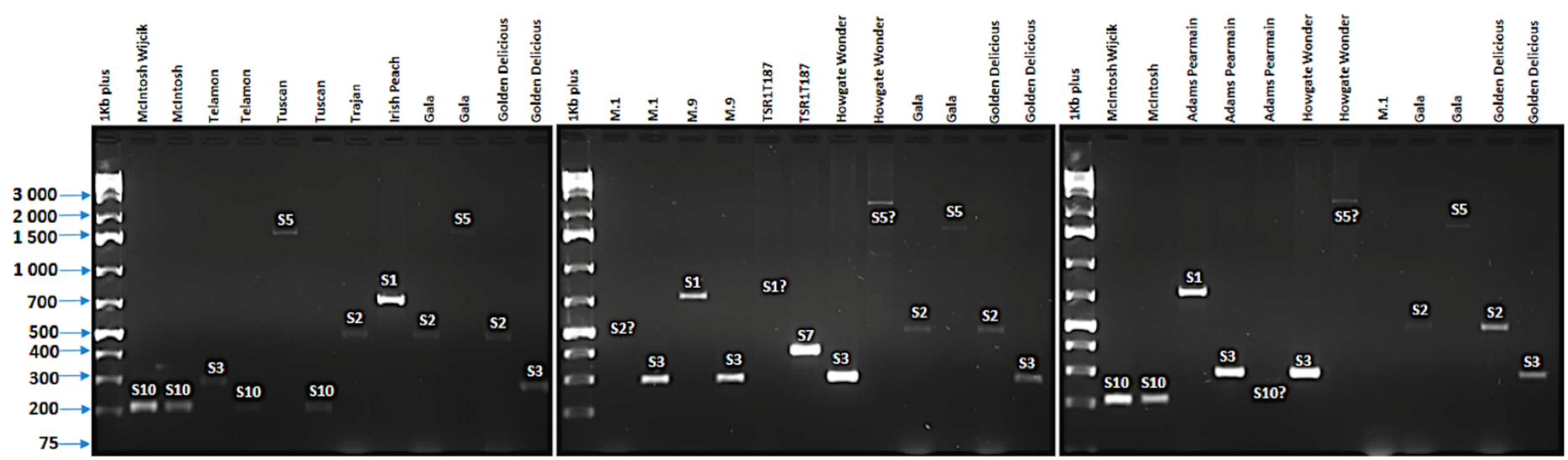


**Figure 3.3. Electrophoretic separation on a 2% agarose gel showing amplified parental S-allele band patterns with allele-specific primers of the apple S-RNase gene.** Each lane shows the outcome of a single allele-specific PCR amplicon as indicated. 1Kb plus denotes a GeneRuler 1 kb plus-DNA ladder used as a molecular size marker estimator in base pairs.

## 3.5 Conclusion

According to the results obtained here and in the following chapter, it was considered that the crinkle dwarf phenotype may not be *S*-linked. Of the 13 progenies studied, seven: 'McIntosh' x 'M.9', 'McIntosh' x 'M.1', 'Trajan' x 'M.9', 'Tuscan' x'M.9', 'McIntosh Wijcik' x TSR1T187, 'Irish Peach' x 'M.9' and TSR1T187 x 'M.9' were fully compatible, and all segregated for crinkled dwarf phenotypes at an approximate ratio of 3:1 or 9:7, though the progeny of 'McIntosh Wijcik' x TSR1T187 did not segregate for crinkled dwarfs in the sec growing season, which may be due to the presence of other modifier genes. The remaining six progenies: 'M.1'x 'M.9', 'Telamon' x 'M.1', 'Telamon' x 'M.9', 'McIntosh' x 'Telamon', 'McIntosh' x 'Tuscan' and 'Irish Peach' x 'M.9' were semi-compatible. The progenies of 'Telamon' x 'M.1'and 'Telamon' x 'M.9', though semi-compatible, sharing an $S_3$ allele segregated unexpectedly for crinkled dwarf phenotypes at approximate 3:1 and 9:7, respectively, thus posing uncertainties on the understanding of the inheritance of the crinkled dwarf trait. The segregation may be due to the fact that 'M.1' and 'M.9' are rootstock cultivars and are not related to 'Telamon', a scion cultivar of 'McIntosh' derivative, and therefore the '*d*' allele could possibly not be linked to $S_3$ in 'Telamon'. However, the progenies of 'M.1' and 'M.9' are both rootstock cultivars sharing the $S_3$ allele, and 'McIntosh' x 'Telamon' and 'McIntosh' x 'Tuscan', are both derivates of 'McIntosh', sharing the $S_{10}$ allele, and assuming 'Irish Peach' and 'M.9' are closely related, exhibited no crinkled dwarf segregation. Therefore, there is a high chance that $S_3$ in 'M.1' and 'M.9'; $S_{10}$ in 'Telamon' and 'Tuscan'; and $S_1$ in 'Irish Peach' and 'M.9' are coupled with the "*d*" allele due to linkage disequilibrium. The power of the *S*-linkage test remains limited. Further work to test all proposed *priori S*-linkage hypotheses, by raising the same crosses and their reciprocals with a larger number of seedlings is needed to further elucidate the molecular genetics underlying the crinkled dwarf trait. Most importantly, due to observed segregations different to what is expected from control by a single recessive gene, a new genetic hypothesis based on a two-gene system needs to be postulated and tested for in additional progenies.

## 3.6 Supplementary Data

Please see back of document

Supplementary Table 3.1 Morphological characteristics of 'McIntosh' x 'M.1' measured between normal and crinkle dwarf seedlings 6 months after germination.

Supplementary Table 3.2. List of consensus and allele-specific primers of the apple S-RNase.

## 3.7 References

**Agapito-Tenfen SZ, Cibele A, Dantas DM, Denardi F, Nodari RO**. 2015. Identification of the *Er 1* resistence gene and *RNase S*-alleles in *Malus prunifolia* var . ringo rootstock. Scientia Agricola **72**, 62–68.

**Alston FH**. 1976. Dwarfing and lethal genes in apple progenies. Euphytica **25**, 505–514.

**Anderson MA, McFadden GI, Bernatzky R, Atkinson A, Orpin T, Dedman H, Tregear G, Fernley R, Clarke AE**. 1989. Sequence variability of three alleles of the self-incompatibility gene of Nicotiana alata. The Plant Cell **1**, 483–491.

**Bai M-Y, Fan M, Oh E, Wang ZY**. 2012. A triple helix-loop-helix/basic helix-loop-helix cascade controls cell elongation downstream of multiple hormonal and environmental signaling pathways in *Arabidopsis*. The Plant Cell **24**, 4917–4929.

**Bošković R, Russell K, Tobutt KR**. 1997. Inheritance of stylar ribonucleases in cherry progenies, and reassignment of incompatibility alleles to two incompatibility groups. Euphytica **95**, 221–228.

**Bošković R, Tobutt KR**. 1996. Correlation of stylar ribonuclease zymograms with incompatibility alleles in sweet cherry. Euphytica **90**, 245–250.

**Broothaerts W, Janssens GA, Proost P, Broekaert WF**. 1995. cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. Plant Molecular Biology **27**, 499–511.

**Broothaerts W, Nerum I Van, Keulemans J**. 2004. Update on and review of the incompatibility (*S*-) genotypes of apple cultivars. HortScience **39**, 943–947.

**Certal AC, Sanchez AM, Kokko H, Broothaerts W, Oliveira MM, Feijó JA**. 1999. S-RNases in apple are expressed in the pistil along the pollen tube growth path. Sexual Plant Reproduction **12**, 94–98.

**Cheng J, Han Z, Xu X, Li T**. 2006. Isolation and identification of the pollen-expressed polymorphic F-box genes linked to the S-locus in apple (*Malus* x *domestica*). Sexual Plant Reproduction **19**, 175–183.

**Claessen H, Keulemans W, Van de Poel B, De Storme N**. 2019. Finding a compatible partner: self-incompatibility in European pear (*Pyrus communis*); molecular control, genetic determination, and impact on fertilization and fruit set. Frontiers in Plant Science **10**.

**Daviere J-M, Achard P**. 2013. Gibberellin signaling in plants. Development **140**, 1147–1151.

**DeNettancourt D**. 1997. Incompatibility in angiosperms. Sexual Plant Reproduction **10**, 185–199.

**DeNettancourt D**. 2001. Incompatibility and incongruity in wild and cultivated plants. Springer-Verlag Berlin. pp. 217-251.

**Doyle J, Doyle J**. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19**, 11–15.

**Dreesen RSG, Vanholme BTM, Luyten K, van Wynsberghe L, Fazio G, Roldán-Ruiz I, Keulemans J**. 2010. Analysis of *Malus S-RNase* gene diversity based on a comparative study of old and modern apple cultivars and European wild apple. Molecular Breeding **26**, 693–709.

**Fazio G, Wan Y, Kviklys D, Romero L, Adams R, Strickland D, Robinson T**. 2014. *Dw2*, a new dwarfing locus in apple rootstocks and its relationship to induction of early bearing in apple scions. Journal of the American Society for Horticultural Science **139**, 87–98.

**Fernández-Fernández F, Evans KM, Clarke JB, Govan CL, James CM, Marič S, Tobutt KR**. 2008. Development of an STS map of an interspecific progeny of *Malus*. Tree Genetics and Genomes **4**, 469–479.

**De Franceschi P, Cova V, Tartarini S, Dondini L**. 2016. Characterization of a new apple *S-RNase* allele and its linkage with the *Rvi5* gene for scab resistance. Molecular Breeding **36**, 1–11.

**Garvey EJ**. 1985. Dwarfism, self-incompatibility, and female sterility in vaccinium ashei (Reade). PhD Dissertation. University of Florida.

**Garvey EJ, Lyrene PM**. 1987. Inheritance of compact growth habit in rabbiteye blueberry. Journal of the American Society for Horticultural Science **112**, 1004–1008.

**Gu C, Wang L, Korban SS, Han Y**. 2015. Identification and characterization of S-RNase genes and S-genotypes in Prunus and Malus species.

**Hafizi A, Shiran B, Maleki B, Imani A, Banović B**. 2013. Identification of new S-RNase self-incompatibility alleles and characterization of natural mutations in Iranian almond cultivars. Trees - Structure and Function **27**, 497–510.

**Harbord RM, Napoli CA, Robbins TP**. 2000. Segregation distortion of T-DNA markers linked to the self- incompatibility (*S*) locus in *Petunia hybrida*. Genetics **154**, 1323–1333.

**Herrera S, Lora J, Hormaza JI, Herrero M, Rodrigo J**. 2018. Optimizing production in the new generation of apricot cultivars: Self-incompatibility, S-RNase allele identification, and incompatibility group assignment. Frontiers in Plant Science **9**, 1–12.

**Hiratsuka S, Zhang SL**. 2002. Relationships between fruit set, pollen-tube growth, and S-RNase concentration in the self-incompatible Japanese pear. Scientia Horticulturae **95**, 309–318.

**House AP**. 2007. Recent findings of the tree fruit self-incompatibility studies. **13**, 7–15.

**Ishimizu T, Endo T, Yamaguchi-Kabata Y, Nakamura KT, Sakiyama F, Norioka S**. 1998. Identification of regions in which positive selection may operate in S-RNase of *Rosaceae*: Implication

for S-allele-specific recognition sites in *S-RNase*. FEBS Letters **440**, 337–342.

**Ishimizu T, Inoue K, Shimonaka M, Saito T, Terai O, Norioka S**. 1999. PCR-based method for identifying the S-genotypes of Japanese pear cultivars. Theoretical and Applied Genetics **98**, 961–967.

**Janssens GA, Goderis IJ, Broekaert WF, Broothaerts W**. 1995. A molecular method for S-allele identification in apple based on allele-specific PCR. Theoretical and Applied Genetics **91**, 691–698.

**Kakui H, Kato M, Ushijima K, Kitaguchi M, Kato S, Sassa H**. 2011. Sequence divergence and loss-of-function phenotypes of S locus F-box brothers genes are consistent with non-self recognition by multiple pollen determinants in self-incompatibility of Japanese pear (*Pyrus pyrifolia*). Plant Journal **68**, 1028–1038.

**Kakui H, Tsuzuki T, Koba T, Sassa H**. 2007. Polymorphism of SFBB -γ And its use for S genotyping in Japanese pear (*Pyrus pyrifolia*). Plant Cell Reports **26**, 1619–1625.

**Kao T, Tsukamoto T**. 2004. The molecular and genetic bases of S-RNase-based self-incompatibility. The Plant Cell **16**, S72–S83.

**Kitahara K, Matsumoto S**. 2002*a*. Sequence of the S10 cDNA from 'McIntosh' apple and a PCR-digestion identification method. HortScience **37**, 187–190.

**Kitahara K, Matsumoto S**. 2002*b*. Cloning of the S25cDNA from 'McIntosh' apple and an S25-allele identification method. Journal of Horticultural Science and Biotechnology **77**, 724–728.

**Kobel F**. 1939. Weitere Untersuchungen uber die Befruchtungsverhaltnisse der Apfelund Birnsorten. Landw Jahrb Schweiz **53**, 160–191.

**Komorisono M**. 2005. Analysis of the rice mutant dwarf and gladius leaf 1. Aberrant Katanin-mediated microtubule organization causes up-regulation of gibberellin biosynthetic genes

independently of gibberellin signaling. Plant Physiology **138**, 1982–1993.

**Long S, Li M, Han Z, Wang K, Li T**. 2010. Characterization of three new S-alleles and development of an S-allele-specific PCR system for rapidly identifying the S-genotype in apple cultivars. Tree Genetics and Genomes **6**, 161–168.

**Maliepaard C, Alston FH, van Arkel G, *et al.*** 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theoretical and Applied Genetics **97**, 60–73.

**Martí A, Alonso JM, Socias I Company R, Wirthensohn M, Hrmova M**. 2011. Molecular modelling of RNases from almond involved in self-incompatibility. Acta Horticulturae.641–644.

**Matsumoto S, Yamada K, Shiratake K, Okada K, Abe K**. 2010. Structural and functional analyses of two new S-RNase alleles, Ssi5 and Sad5, in apple. Journal of Horticultural Science and Biotechnology **85**, 131–136.

**McClure B**. 2009. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for. Journal of Experimental Botany, **60**, 1069–1081.

**McClure B, Franklin-Tong V**. 2006. Gametophytic self-incompatibility : understanding the cellular mechanisms involved in "self " pollen tube inhibition. Planta **224**, 233–245.

**McClure BA, Gray JE, Anderson MA, Clarke AE**. 1990. Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature **347**, 757–760.

**Milach SCK, Federizzi LC**. 2001. Dwarfing genes in plant improvement. Advances in Agronomy **73**, 35–63.

**Minamikawa M, Kakui H, Wang S, Kotoda N, Kikuchi S, Koba T, Sassa H**. 2010. Apple S locus region represents a large cluster of related, polymorphic and pollen-specific F-box genes. Plant

Molecular Biology **74**, 143–154.

**Van Nerum I, Geerts M, Van Haute A, Keulemans J, Broothaerts W**. 2001. Re-examination of the self-incompatibility genotype of apple cultivars containing putative 'new' S-alleles. Theoretical and Applied Genetics **103**, 584–591.

**Newbigin E, Anderson MA, Clarke AE**. 1993. Gametophytic self-incompatibility systems. The Plant Cell **5**, 1315–1324.

**Newbigin E, Uyenoyama MK**. 2005. The evolutionary dynamics of self-incompatibility systems. Trends in Genetics **21**, 500–505.

**O'Leary MC, Boyle TH**. 1998. Segregation distortion at isozyme locus *Lap-1* in Schlumbergera (*Cactaceae*) is caused by linkage with the gametophytic self-incompatibility (S) locus. Journal of Heredity **89**, 206–210.

**Okada K, Tonaka N, Taguchi T, Ichikawa T, Sawamura Y, Nakanishi T, Takasaki-Yasuda T**. 2011. Related polymorphic F-box protein genes between haplotypes clustering in the BAC contig sequences around the S-RNase of Japanese pear. Journal of Experimental Botany **62**, 1887–1902.

**Ortega E, Bošković RI, Sargent DJ, Tobutt KR**. 2006. Analysis of S-RNase alleles of almond (*Prunus dulcis*): Characterization of new sequences, resolution of synonyms and evidence of intragenic recombination. Molecular Genetics and Genomics **276**, 413–426.

**Potter D, Eriksson T, Evans RC, *et al.*** 2007. Phylogeny and classification of *Rosaceae*. Plant Systematics and Evolution **266**, 5–43.

**Pratas MI, Aguiar B, Vieira J, Nunes V, Teixeira V, Fonseca NA, Iezzoni A, Van Nocker S, Vieira CP**. 2018. Inferences on specificity recognition at the *Malus domestica* gametophytic self-incompatibility system. Scientific Reports **8**, 1–17.

**Rabbani T, Mohammad, Mirlohi A, Saeidi G, R. Sabzalian M**. 2012. An evaluation of segregation distortion in wide crosses of safflower. International Journal of Agriculture and Forestry **2**, 288–293.

**Richman A**. 2000. Evolution of balanced genetic polymorphism. Molecular Ecology **9**, 1953–1963.

**Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T**. 2007. S locus F-box brothers: Multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics **175**, 1869–1881.

**Sassa H, Koba T, Ikehashi H, Nishio T, Kowyama Y, Hirano H**. 1996. Self-incompatibility (S) alleles of the *Rosaceae* encode members of a distinct class of the T2/S ribonuclease superfamily. MGG Molecular & General Genetics **250**, 547–557.

**Sato Y, Sentoku N, Miura Y, Hirochika H, Kitano H, Matsuoka M**. 1999. Loss-of-function mutations in the rice homeobox gene OSH15 affect the architecture of internodes resulting in dwarf plants. EMBO Journal **18**, 992–1002.

**Seleznyova AN, Tustin DS, Thorp TG**. 2008. Apple dwarfing rootstocks and interstocks affect the type of growth units produced during the annual growth cycle: Precocious transition to flowering affects the composition and vigour of annual shoots. Annals of Botany **101**, 679–687.

**Sonneveld T, Robbins TP, Bošković R, Tobutt KR**. 2001. Cloning of six cherry self-incompatibility alleles and development of allele-specific PCR detection. Theoretical and Applied Genetics **102**, 1046–1055.

**Tao R, Habu T, Yamane H, Sugiura A, Iwamoto K**. 2000. Molecular markers for self-compatibility in Japanese apricot (*Prunus mume*). HortScience **35**, 1121–1123.

**Tao R, Yamane H, Sassa H, Mori H, Gradziel TM, Dandekar AM, Sugiura A**. 1997. Identification of stylar RNases associated with gametophytic self-incompatibility in almond (*Prunus dulcis*). Plant

and Cell Physiology **38**, 304–311.

**Tao R, Yamane H, Sugiura A, Murayama H, Sassa H, Mori H**. 1999. Molecular typing of S-alleles through identification, characterization and cDNA cloning for S-RNases in sweet cherry. Journal of the American Society for Horticultural Science **124**, 224–233.

**Tong H, Liu L, Jin Y, Du L, Yin Y, Qian Q, Zhu L, Chu C**. 2012. Dwarf and low-tillering acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate Brassinosteroid responses in rice. The Plant Cell **24**, 2562–2577.

**Tworkoski T, Miller S**. 2007. Rootstock effect on growth of apple scions with different growth habits. Scientia Horticulturae **111**, 335–343.

**Ushijima K, Sassa H, Tamura M, Kusaba M, Tao R, Gradziel TM, Dandekar AM, Hirano H**. 2001. Characterization of the S-locus region of almond (*Prunus dulcis*): Analysis of a somaclonal mutant and a cosmid contig for an S haplotype. Genetics **158**, 379–386.

**Verdoodt L, Van Haute A, Goderis IJ, De Witte K, Keulemans J, Broothaerts W**. 1998. Use of the multi-allelic self-incompatibility gene in apple to assess homozygocity in shoots obtained through haploid induction. Theoretical and Applied Genetics **96**, 294–300.

**Wu J, Gu C, Khan MA, Wu J, Gao Y, Wang C, Korban SS, Zhang S**. 2013. Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of *Rosaceae*. Critical Reviews in Plant Sciences **32**, 53–68.

**Wünsch A, Hormaza JI**. 2004. S-allele identification by PCR analysis in sweet cherry cultivars. Plant Breeding **123**, 327–331.

**Yang W, Ren S, Zhang X, *et al.*** 2011. Bent uppermost internode1 encodes the class II formin FH5 crucial for actin organization and rice development. The Plant Cell **23**, 661–680.

**Zamir D, Tadmor Y**. 1986. Unequal segregation of nuclear genes in plants. Botanical Gazette **147**, 355–358.

**Zuccherelli S, Tassinari P, Broothaerts W, Tartarini S, Dondini L, Sansavini S**. 2002. S-allele characterization in self-incompatible pear (*Pyrus communis* L.). Sexual Plant Reproduction **15**, 153–158.

# Chapter 4
# Molecular characterisation and mapping of genes associated with crinkle dwarf trait in apple

This chapter is prepared in accordance to the style of the journal Molecular Breeding, but it is more elaborate for the purpose of the thesis.

Z.T.L. Mbulawa[1,2], A.E. Bester-van der Merwe[2,] J. Kriel[1], M. Troggio[3], M.K. Soeker[1], and K.R. Tobutt[1]

*[1]ARC Infruitec-Nietvoorbij, Private Bag X5026, Stellenbosch, 7599, South Africa*

*[2]Department of Genetics, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa*

*[3]Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy*

## 4.1. Abstract

Crinkle dwarfs, a type of hybrid necrosis, are considered economically unfavourable due to their undesirable characteristics for fruit production including a typical dwarf seedling phenotype associated with crinkled leaves, poor growth, and in some cases lethality. Little is known about the precise physiological or genetic mechanism controlling the crinkle dwarf phenotype. In this study, a mapping population of 94 F1 progeny from 'McIntosh' and 'M.1', and segregated 3:1 for normal *versus* crinkle dwarf ($\chi^2 = 0.10$, $p = 0.750$, n.s.), a homozygous recessive trait, denoted as *crinkledw* was attributed to control by a single recessive gene.

High-density genetic maps for the two parents as well as the consensus map were constructed using the apple 20K Infinium® single nucleotide polymorphisms (SNP) array. The 'McIntosh' (maternal parent) map spanned 1171.96 cM, with 3190 SNPs at an average inter-marker spacing of 0.40 cM. The total length of the 'Malling 1' ('M.1') (paternal parent) map was 1402.44 cM, with 2640 SNPs at an average

inter-marker distance of 0.56 cM. The *crinkledw* trait was mapped on linkage group (LG)8 in the 'McIntosh' map at position 54.38 cM, but on LG2 in 'M.1' at position 5.63 cM.

Additionally, *crinkledw* trait was validated using a Kruskal-Wallis (KW) non-parametic approach. The significance tests for KW also indicated the trait to be on LG8 in 'McIntosh'. The highest K* score of 30.67 ($p < 0.0001$), declared a significant marker-trait association at a genetic position of 54.28 cM. The *crinkledw* obtained from KW localized in close-proximity with the mapped "*crinkledw1"* (54.38 cM) at differing genetic distances of 0.078 cM and 0.922 cM, respectively, with marker SNP_FB_0765586 tightly linked with the highest K* score. Based on the consensus genetic map, *crinkledw* co-segregated with GDsnp02575 also on LG8, though high segregation distortion was observed. To our knowledge, this is the first study to map and identify a SNP marker associated with crinkle dwarf phenotype in apple. A marker/probe that is specific for crinkle dwarf trait could be designed, and used to screen other mapping progenies segregating for crinkle dwarf.

## 4.2. Introduction

The domesticated apple (*Malus pumila* Mill.) ($2n=2x=34$), is one of the most economically important fruit crops worldwide (Fernández-Fernández *et al.*, 2014). It belongs to the *Spireaoidea*e subfamily of *Rosaceae*, under the genus *Malus* together with pears (*Pyrus* spp.) and other cultivated tree species such as quince (*Cydonia oblonga*), loquat (*Eriobotryajaponica*). and medlar (*Mespilus germanica*) (Potter *et al.*, 2007). Cultural practices for controlling growth and vigour in commercial apple production relies primarily on the use of dwarfing rootstocks, mainly 'M.9', 'M.26' and 'M.27'(Atkinson and Else, 2001; Costes and García-Villanueva, 2007). The control of tree size is of utmost importance for the optimisation of fruit productivity and efficiency of orchard breeding strategies (Webster, 1995; Fallahi *et al.*, 2002; Byrne, 2012). Accordingly, two major QTLs: *Dw1* and *Dw*2, conferred to rootstock-induced dwarfing of apple scions, have been identified on LG5 and LG11 of 'M.9', respectively (Pilcher *et al.*, 2008; Fazio *et al.*, 2014; Foster *et al.*, 2015; Harrison *et al.*, 2016).

Alston (1976) carried out a study on the genetics of three rare forms of dwarf types in apple, one, crinkle dwarf, was reported in a cross of 'Irish Peach' x TSR1T187 which segregated 3:1 for normal

*versus* crinkle dwarf. The crinkled phenotype was attributed to control by a single recessive gene, where both parents are heterozygous, and assigned a gene symbol *cr*. Crinkle dwarf phenotypes are characterised by small rounded crinkled leaves with normal internodes (Alston, 1976). This trait has not been mapped. This phenotype may be similar to the one investigated in the current study.

Crinkle dwarf phenotype could be considered a type of hybrid incompatibility (hybrid necrosis), a form of postzygotic reproductive isolation, which occurs in the seedling or adult stage and is often associated with symptoms such as wilting, chlorosis, stunted growth, and lethality (Bomblies and Weigel, 2007; Chen *et al.*, 2016). In recent years, hybrid incompatibility have been reported in interspecific and intraspecific hybrids including *Arabidopsis* (Bomblies *et al.*, 2007), wheat (*Triticum* spp.) (Takumi *et al.*, 2013) , tobacco (*Nicotiana* spp.) (Mizuno *et al.*, 2010), and recently in pear *Pyrus* spp. (Montanari *et al.*, 2016). Hybrid incompatibility remains a serious threat in the agricultural and breeding sectors (Orr and Presgraves, 2000).

A key resource in support of classical genetics and genomics is the construction of dense genetic linkage maps (Liebhard *et al.*, 2003; Antanaviciute *et al.*, 2012; Troggio *et al.*, 2012; Wang *et al.*, 2016). This serves as a prerequisite for studying the inheritance of both qualitative and quantitative traits, thus facilitating the integration of molecular markers into marker-assisted selection (MAS) (Patocchi *et al.*, 2009; Shiratake and Suzuki, 2016; Nadeem *et al.*, 2018). Single nucleotide polymorphisms (SNPs), are the most abundant type of molecular markers with a low but stable mutation rate across the genome (Chagné *et al.*, 2008; Lateef, 2015; Huq *et al.*, 2016).

The development and application of high-throughput SNP arrays has gained remarkable attention in recent years, enabling simultaneous screening of thousands of polymorphic loci at a lower cost per data point in comparison to more traditional marker technologies such as microsatellite genotyping (Rafalski, 2002; Ganal *et al.*, 2012; Rasheed *et al.*, 2017; Nadeem *et al.*, 2018). Currently, three apple SNP arrays have been developed, i.e. the International RosBREED SNP Consortium (IRSC) apple 8K, FruitBreedomics 20K Illumina Infinium®, and Axiom® 480K (Chagné *et al.*, 2012; Bianco *et al.*, 2014, 2016). Recently, a newly developed, apple International RosBREED SNP Consortium OpenArray v1.0 (IRSCOA v1.0) has been available (Chagné *et al*., 2019).

To date, numerous SNP arrays have been developed in other *Rosaceae* species e.g. the 9K peach array (Verde *et al.*, 2012), 90K Axiom® array for strawberry (Bassil *et al.*, 2015), RosBREED 6K Illumina Infinium® cherry SNP array (Peace *et al.*, 2012). In pear, 1K (Montanari *et al.*, 2013) and two newly developed Axiom™ 70K SNP (Montanari *et al.*, 2019) and 200K Axiom® PyrSNP (Li *et al.*, 2019) arrays have been developed. The Axiom® apple 480K array remains the largest in the fruit tree species (Bianco *et al.*, 2016). These arrays have been used for the generation of linkage maps (Antanaviciute *et al.*, 2012; Klagges *et al.*, 2013; Montanari *et al.*, 2013; Clark *et al.*, 2014; Frett *et al.*, 2014; Wu *et al.*, 2014), evaluation of the quality of physical maps (Troggio *et al.*, 2013; Di Pierro *et al.*, 2016), fine mapping and validation of quantitative trait loci (QTL) (Chagné *et al.*, 2019; Peace *et al.*, 2019), elucidation of marker-trait associations (Eduardo *et al.*, 2013; Font i Forcada *et al.*, 2019), genome-wide association studies (Kumar *et al.*, 2013), genomic selection studies (Kumar *et al.*, 2012), validation of pedigrees and verification of trueness-to-type of breeding lines and accessions (Pikunova *et al.*, 2014).

The aim of this chapter is to elucidate the underlying genetics of crinkle dwarf growth habit by constructing high-density genetic linkage maps using the apple 20K Infinium® SNP array (Bianco *et al.*, 2014), with the view to characterize and map the molecular marker(s) that co-segregate with the genomic region associated with crinkle dwarf trait in a progeny of ‘McIntosh’ x ‘M.1’. Marker assisted selection using these linked markers could aid in avoiding raising progenies segregating for this deleterious trait in apple breeding programmes.

## 4.3. Materials and Methods

### 4.3.1 Plant Material

An F1 population of 118 seedlings derived from the cross between ‘McIntosh’ × ‘Malling 1’ (‘M.1’), both of normal habit, was raised with the view to study y to investigate the molecular genetics of dwarf seedlings associated with crinkle dwarf phenotype. ‘McIntosh’, the maternal parent, is a scion cultivar of unknown origin. It is known to carry dwarf genes as indicated in the introduction. It is known to carry dwarf genes. ‘M.1’ is a rootstock cultivar, a parent of the invigorating rootstock, ‘M.M.106’

('M.M.106' = 'M.1' x 'Northern Spy'). The seedlings were raised in 2016 and grown on their own roots under natural photoperiodic conditions in the glasshouse at the Agricultural Research Council (ARC) Infruitec-Nietvoobij' Bien Donné Research Farm, Groot Drakenstein, South Africa [(33°83'33"32.06 (S); 18°98'33"33.59 (E)]. To avoid any horticultural influence on tree shape and vigour, the seedlings were not pruned.

The population was visually scored in the first two years, 2016 and 2017, as either "normal", having standard height and healthy leaves, or "crinkle dwarf", with seedlings of short stature with abnormal dark green crinkled leaves, usually brittle in texture.

For mapping, 94 seedlings were selected from the progeny of 118 individuals. Approximately 300 mg of young leaf material, from each seedling and both parents were collected and stored in a 96-deep well plate and freeze-dried for 12 hr. Thereafter, they were shipped to Fondazione Edmund Mach (FEM), San Michele all'Adige, Italy, where genomic DNA extraction and SNP genotyping were conducted.

### 4.3.2 DNA extraction and quantification

Genomic DNA was extracted using the QIAGEN DNeasy® 96 Plant Extraction Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. The resulting genomic DNA was quantified using a NanoDrop™ 2000c spectrophotometer (Thermo Fisher Scientific Inc.). The concentrations of all DNA samples were adjusted to 50 ng/uL for subsequent SNP genotyping.

### 4.3.3 SNP genotyping and data scoring

The population was genotyped with the apple 20K Infinium® SNP chip array (Bianco *et al.*, 2014), following the standard Illumina protocols described by Antanaviciute et al. (2012) and Chagné et al (2012). The raw genotypic iScan data output obtained was imported into the GenomeStudio Genotyping Module software v2.0 (Illumina, Inc., San Diego, California, USA) for analysis.

The SNP genotype calls were filtered through the ASSIsT software (Di Guardo *et al.*, 2015), a filtering and calling pipeline that accounts for the presence of null alleles and signal intensity differences

among A/B genotypes, thus increasing the number of usable SNP markers. SNP loci were further eliminated if missing parental genotypes could not be positively determined on the basis of progeny segregation. The SNP markers with more than 10% missing data were excluded. The individual seedlings that were not consistent with the parental genotypes were removed, as they were expected to be outcrosses, or contaminated samples, in this study, two seedlings were eliminated. The failed and monomorphic markers were excluded, whereas the polymorphic SNPs were further inspected for clustering analysis.

### 4.3.4 Individual genetic linkage map construction, consensus genetic linkage map integration and co-segregation analysis

The analysis of single locus segregations and construction of parental linkage maps was carried out using JoinMap® v5.0 adopting the 'two-way pseudo-testcross' mapping strategy using the cross-pollinated (CP) population option (Grattapaglia and Sederoff, 1994). Cross-pollinated population indicates a cross between two heterozygous diploid parents, with linkage phases originally unknown (Van Ooijen and Voorrips, 2017). The SNP markers were re-coded in GenomeStudio into JoinMap genotype codes for linkage analysis. The ABxAA or ABxBB (segregating in the female 'McIntosh' parent) were coded as <lmxll>, and AAxAB or BBxAB (segregating in the male 'M.1' parent) were coded as <nnxnp>. The SNP markers that segregated ABxAB in both parental genotypes were coded as <hkxhk>.

Due to the recessive inheritance of the crinkle dwarf trait, the chromosomal region containing the mutation is expected to be heterozygous, therefore assigned to a segregation type <hkxhk>. The parental maps were constructed using only SNP markers segregating in one of the parents i.e. segregation types <lmxll> and <nnxnp> with the inclusion of <hkxhk> for crinkle dwarf trait, and were analysed independently to construct separate parental genetic linkage maps. The mapping strategy used in the construction of the parental maps assigned the genotypes of crinkle dwarf trait into two categories *crinkledw1* and *crinkledw2* i.e. "hhxhk" or "hkxkk", respectively.

The consensus genetic linkage map was constructed by utilising the SNP marker types from both the 'McIntosh' and 'M.1' i.e. <lmxll>, <nnxnp>, together with the <hkxhk> which served as bridge markers to integrate the individual parental maps resulting in a single consensus map.

The linkage analysis was performed using the independence threshold grouping parameter with a stringent logarithm of odds (LOD) score of 15 and the Maximum Likelihood (ML) mapping algorithm with maximum recombination frequency of 0.4 (Jansen *et al.*, 2001; Van Ooijen, 2011). A Chi-square ($\chi 2$) goodness-of-fit test ($p < 0.05$) was performed on the segregation data of all markers, and the markers deviating significantly from the expected 1:1 or 1:2:1 ratios were excluded from further analysis. The contribution of each SNP marker to the average goodness-of-fit (mean Chi-square) and the nearest-neighbour fit (N.N. Fit) value was checked to confirm its most likely position in each linkage group. Subsequently, SNP markers at fixed distances and those that were evenly distributed to cover the 17 LGs were used to build framework parental maps. The SNP markers that mapped to the same location, were grouped into single bins with the purpose of reducing map complexity for linkage analysis. A single SNP containing no missing data for a progeny was used for linkage analysis from each bin. Map distances were converted to centiMorgans (cM) using the Kosambi mapping function (Kosambi, 1944). The LGs were numbered according to internationally acknowledged apple genomes (Velasco *et al.*, 2010; Daccord *et al.*, 2017). Graphical presentations of genetic linkage maps were generated using MapChart 2.2 software (Voorrips, 2002).

### 4.3.5 Comparison of parental genetic maps to physical positions

To evaluate the quality of the parental genetic linkage map, the consistency of locus (marker) order were compared against the reference genome. The genetic distances of the parental maps were aligned with their physical position on the 'Golden Delicious' double haploid apple reference genome (GDDH13) v1.1 (Daccord *et al.*, 2017). The visualisation of recombination along the linkage groups were plotted using MareyMap package (Rezvoy *et al.*, 2007). The estimated recombination rate (cM/Mb) between each pair of adjacent markers were calculated as the ratio of genetic distance in cM against physical distance in Mb.

### 4.3.6 Validation of crinkle dwarf trait using Kruskal-Wallis

The co-segregation analysis and mapping of the *crinkledw1* and *crinkledw2* were validated using the non-parametric Kruskal-Wallis (KW) test (Kruskal and Wallis, 1952). The KW test, a non-parametric method is mostly used to detect association-marker-traits, for which markers at K* value at $p < 0.001$ are considered significant (Kruglyak and Lander, 1995; Kruglyak *et al.*, 1996; Rebai, 1997). The KW analysis was conducted in MapQTL® v6.0 (Van Ooijen, 2009).

## 4.4 Results

### 4.4.1 Inheritance and phenotypic analysis of crinkle dwarf

Phenotypically, dwarf seedlings were short in stature and typically had dark green crinkled leaves, and could clearly be distinguished from normal seedlings by about three months (Figure 4.1). The 118 seedlings in the full progeny segregated 90:28 normal *versus* crinkle dwarf, approximating to 3:1 ($\chi^2$ = 0.10, $p$ = 0.75, n.s.). The results are consistent with segregation of a single gene, for which the two parents, 'McIntosh' (*DdEe*) and 'M.1' (*DDEe*), are heterozygous (Table 4.1). Therefore, in this progeny the crinkle dwarf phenotype is expressed when one of the genes is homozygous recessive i.e. *D-ee*.

**Table 4.1. Phenotypic segregation of crinkle dwarf on the F1 population of 'McIntosh' x 'M.1'.**

| Maternal | Paternal | Normal seedlings | Crinkled dwarf seedlings | Expected segregation | $\chi^2$ | $p$-value |
|---|---|---|---|---|---|---|
| McIntosh ((*DdEe*) | M1 (*DDEe*) | 90 | 28 | 3:1 | 0.10 | 0.750 |

Chi-square ($\chi^2$, $p < 0.05$, df = 1) were calculated under the assumption of a Mendelian 3:1 segregation ratio

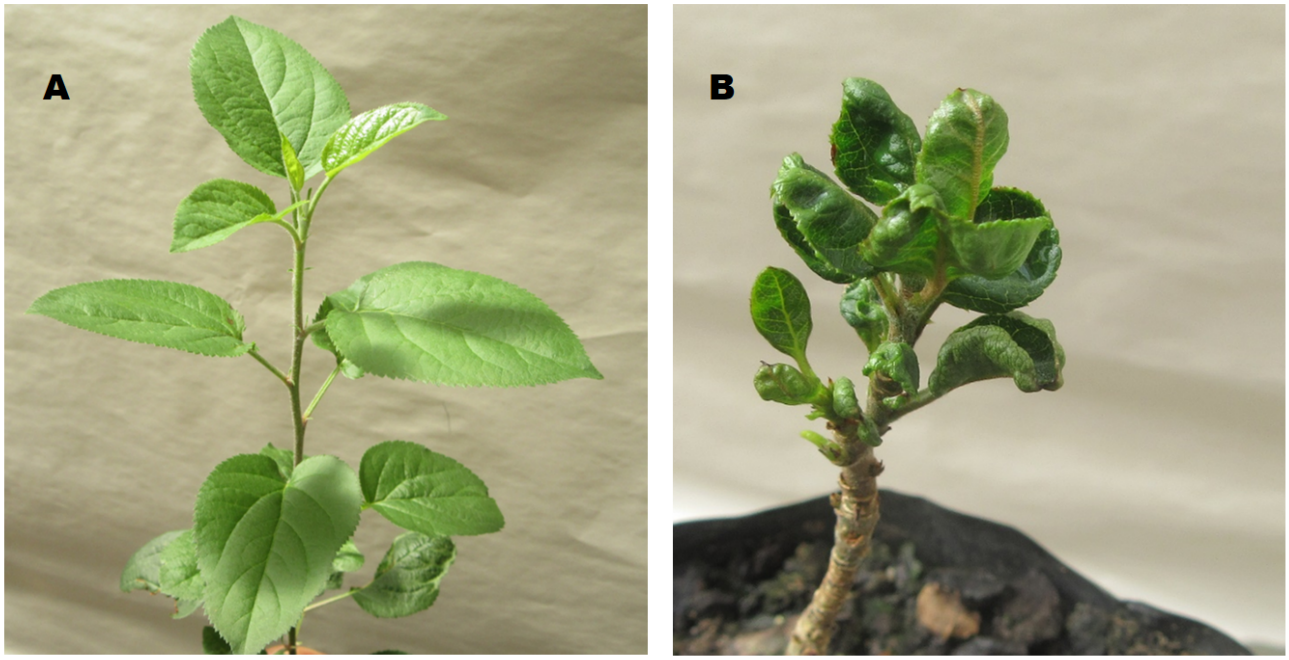


**Figure 4.1. Distinct phenotypes between normal (A) and dwarf seedlings showing dark-green, crinkled leaves (B) in a F1 mapping population of 'McIntosh' x 'M.1'.**

### 4.4.2 SNP genotyping

GenTrain scores for all SNPs generated for the progeny of 'McIntosh' and 'M.1'ranged from 0.0059 to 0.916, with an average of 0.734. Cluster separation ranged from 0.0306 to 1 with an average of 0.828. Of the 18,019 SNP markers included on the 20K SNP array, 9,961 (55.3 %) were discarded, of which 7,157 (39.7 %) were monomorphic, 2,310 (12.8%) failed to amplify, 479 were distorted with unexpected segregation and 15 (0.1 %) SNPs showed null alleles. The remaining 8,058 SNP markers, 7,734 (42.9%) were considered polymorphic and further used in the construction of genetic linkage maps. The statistical performance of the 20K SNP array is summarized (Table 4.2). Prior to linkage analysis two individual seedlings (Normal21 and Normal27) were discarded due to being identified as outcrosses.

**Table 4.2. Summary statistics for the 20K Infinium® SNP array applied in genotyping the progeny 'McIntosh' x 'M.1'.**

| SNPs Parameters | No. of SNPs | Percentage (%) |
|---|---|---|
| Approved | 8058 | 44.7 |
| Monomorphic | 7157 | 39.7 |
| Failed | 2310 | 12.8 |
| Distorted segregation | 479 | 2.7 |
| Failed null alleles | 15 | 0.1 |
| Total | 18019 | 100 |

### 4.4.3 SNP segregation analysis for the 20K Infinium® SNP array applied to the progeny 'McIntosh' x 'M.1'

Of the 8058 polymorphic SNPs, four segregation types were observed: 3289 <lmxll>, 2718 <nnxnp>, 1943 <hkxhk> and 106 were <efxeg>. Upon conducting linkage analysis using different segregation combinations (data not shown), inconsistency were observed in the generated maps. Therefore, only the SNP markers with <lmxll> and <nnxnp> segregations were used for the construction of the parental maps, while the marker types <lmxll>, <nnxnp> and <hkxhk> were included for the consensus map.

### 4.4.4 Parental genetic linkage maps

The genetic linkage maps of the two parents, 'McIntosh' and 'M.1' were constructed separately and spanned across 17 LGs representing the number of chromosomes in the haploid apple genome. The maternal 'McIntosh' genetic map consisted of 3190 SNP loci, which spanned a total map length of 1171.96 cM. This resulted in an average marker spacing of 0.40 cM, with only two gaps exceeding 10 cM. The latter were observed between adjacent markers on LG8 and LG9 at 13.66 cM and 15.11 cM, respectively. The number of SNP markers per chromosome ranged from 83 in LG16 to 353 in LG15.

The longest linkage group was LG15 and measured 111.30 cM with the highest number of SNP loci (353), with LG14 consisting of 125 SNPs and being the shortest (53.72 cM) (Table 4.3; Figure 4.2).

The paternal 'M.1' genetic map consisted of 2640 SNP markers, which spanned a total length of 1402.44 cM. The average marker spacing was 0.56 cM with nine gaps exceeding 10 cM and observed between adjacent markers in LGs 13, 10, 5, 7, 12, 6, 4, 2, 15 (sequential order of the greatest distance) at 21.67 cM, 18.39 cM, 16.82 cM, 15.31 cM, 15.31 cM, 13.84 cM, 12.41 cM, 11.02 cM, and 11.02 cM, respectively. The number of SNP markers per chromosome ranged from 99 in LG7 to 235 in LG16. LG15 was the longest linkage group and measured 134.31 cM and consisting of 226 SNP markers. LG8 was the shortest at 45.13 cM consisting of 123 SNP markers. It was also observed that LG16 consisted of the highest number of SNPs (235) but only measured 54.07 cM, which was contrary to that of 'McIntosh' in which the greatest number of SNPs (353) resulted in the longest linkage group, LG15 (Table 4.3; Figure 4.2). It is important to note that the individual parental maps could not be aligned nor integrated into a consensus map, as common and bridge SNPs and additional tested SSRs created tension in the formation of the grouping order, excluding other markers, leading to spurious linkage results and incorrect orders.

***Co-segregation analysis and mapping of crinkle dwarf trait***

The two assigned trait categories *crinkledw1* and *crinkledw2* mapped on LG8 of the 'McIntosh' linkage map, and the linkage group contained 228 SNP markers. The *crinkledw1* and *crinkledw2* segregated in a 1:3 (hh:hk) and 3:1(hk:kk) Mendelian ratio in accordance with monogenic trait segregation, respectively, with the rest of other SNPs segregating 1:1 (ll:lm) (Supplementary Table 4.3).

On the 'M.1' genetic map, *crinkledw1* and *crinkledw2* mapped on LG2, and the linkage group consisted of 168 SNP markers. The *crinkledw1* and *crinkledw2* segregated in a 1:3 (hh:hk) and 3:1(hk:kk) Mendelian ratio in accordance with monogenic trait segregation , respectively, with the rest of other SNPs segregating 1:1 (nn:np) (Supplementary Table 4.4). The detailed maps of LG2 and LG8 for both parents are depicted in Figure 4.3.

In the 'McIntosh' genetic linkage map, the *crinkledw1* was located at genetic position 54.38 cM co-segregating with seven other markers: RosBREEDSNP_SNP_CT_32814468, SNP_FB_0765586, SNP_FB_0765594, SNP_FB_0461119, SNP_FB_0461117, SNP_FB_1053497 and SNP_FB_0765947. This mapped position was flanked by markers: GDsnp00975, SNP_FB_0765104 and SNP_FB_0765111 at 53.23 cM, and, markers: SNP_FB_0811059, SNP_FB_0811041 and SNP_FB_0811048 at 57.77 cM. In contrast, the *crinkledw2* mapped at genetic position of 49.83 cM collating with two other SNP markers: SNP_FB_1034394 and SNP_FB_0762865. The *crinkledw2* region was flanked by GDsnp02575 at 48.72 cM and SNP_FB_1103570 at 50.96 cM The *crinkledw1* and *crinkledw2* are located 4.52 cM apart on LG8 (Supplementary Table 4.3). There were no *crinkledw1* and *crinkledw2* mapped on LG8 in the 'M.1' genetic linkage map.

*Crinkledw1* and *crinkledw2* however did not map on LG8 of the paternal 'M.1' map but on LG2. The *crinkledw traits* both mapped at position 5.63 cM of LG2 co-segregating with SNP_FB_0949262. This region was flanked by SNP_ FB_0451368 and SNP_FB_0452379 at positions 4.522 cM and 6.744 cM, respectively (Supplementary Table 4.4). There were no *crinkledw1* and *crinkledw2* mapped on LG2 in 'McIntosh' genetic linkage map.

**Table 4.3. Distribution of SNP markers across the 17 linkage groups in the constructed parental genetic maps of 'McIntosh' and 'M.1'.**

| Linkage group | McIntosh parent (♀) | | | | M.1 parent (♂) | | | |
|---|---|---|---|---|---|---|---|---|
| | [a]No. of SNPs mapped | Genetic length (cM) | [b]Avg. marker density (cM) | Large gap (cM) | [a]No. of SNPs mapped | Genetic length (cM) | [b]Avg. marker distance (cM) | Large gap (cM) |
| LG 1 | 105 | 60.59 | 0.58 | 5.75 | 154 | 63.69 | 0.41 | 5.82 |
| LG 2 | 260 | 61.96 | 0.24 | 2.22 | 168 | 82.69 | 0.49 | 11.02 |
| LG 3 | 170 | 68.29 | 0.40 | 6.70 | 200 | 93.63 | 0.47 | 9.67 |
| LG 4 | 158 | 54.35 | 0.34 | 3.37 | 151 | 80.11 | 0.53 | 12.41 |
| LG 5 | 225 | 86.81 | 0.39 | 5.75 | 199 | 101.02 | 0.51 | 16.82 |
| LG 6 | 121 | 61.33 | 0.51 | 5.75 | 114 | 78.67 | 0.69 | 13.84 |
| LG 7 | 177 | 64.88 | 0.37 | 5.75 | 99 | 91.20 | 0.92 | 15.31 |
| LG 8 | 228 | 64.82 | 0.28 | 13.66 | 123 | 45.13 | 0.37 | 7.07 |
| LG 9 | 167 | 69.75 | 0.42 | 15.11 | 181 | 87.15 | 0.48 | 8.35 |
| LG 10 | 204 | 71.04 | 0.35 | 5.75 | 110 | 109.27 | 0.99 | 18.39 |
| LG 11 | 221 | 80.02 | 0.36 | 3.37 | 156 | 87.16 | 0.56 | 7.071 |
| LG 12 | 150 | 71.84 | 0.48 | 4.55 | 163 | 89.32 | 0.55 | 15.31 |
| LG 13 | 188 | 63.59 | 0.34 | 3.37 | 124 | 77.53 | 0.62 | 21.67 |
| LG 14 | 125 | 53.71 | 0.43 | 5.75 | 135 | 49.32 | 0.36 | 3.411 |
| LG 15 | 353 | 111.30 | 0.31 | 3.37 | 226 | 134.31 | 0.59 | 11.02 |
| LG 16 | 83 | 54.25 | 0.65 | 8.25 | 235 | 54.07 | 0.23 | 7.07 |
| LG 17 | 255 | 73.41 | 0.29 | 3.37 | 102 | 78.18 | 0.77 | 5.82 |
| **Min SNPs/LG** | **83** | **53.72** | **0.24** | **2.22** | **99** | **45.13** | **0.23** | **3.41** |
| **Max SNPs/LG** | **353** | **111.30** | **0.65** | **15.11** | **235** | **134.31** | **0.99** | **21.67** |
| **Total** | **3190** | **1171.96** | **0.40** | **[c]n.a** | **2640** | **1402.44** | **0.562** | **[c]n.a** |

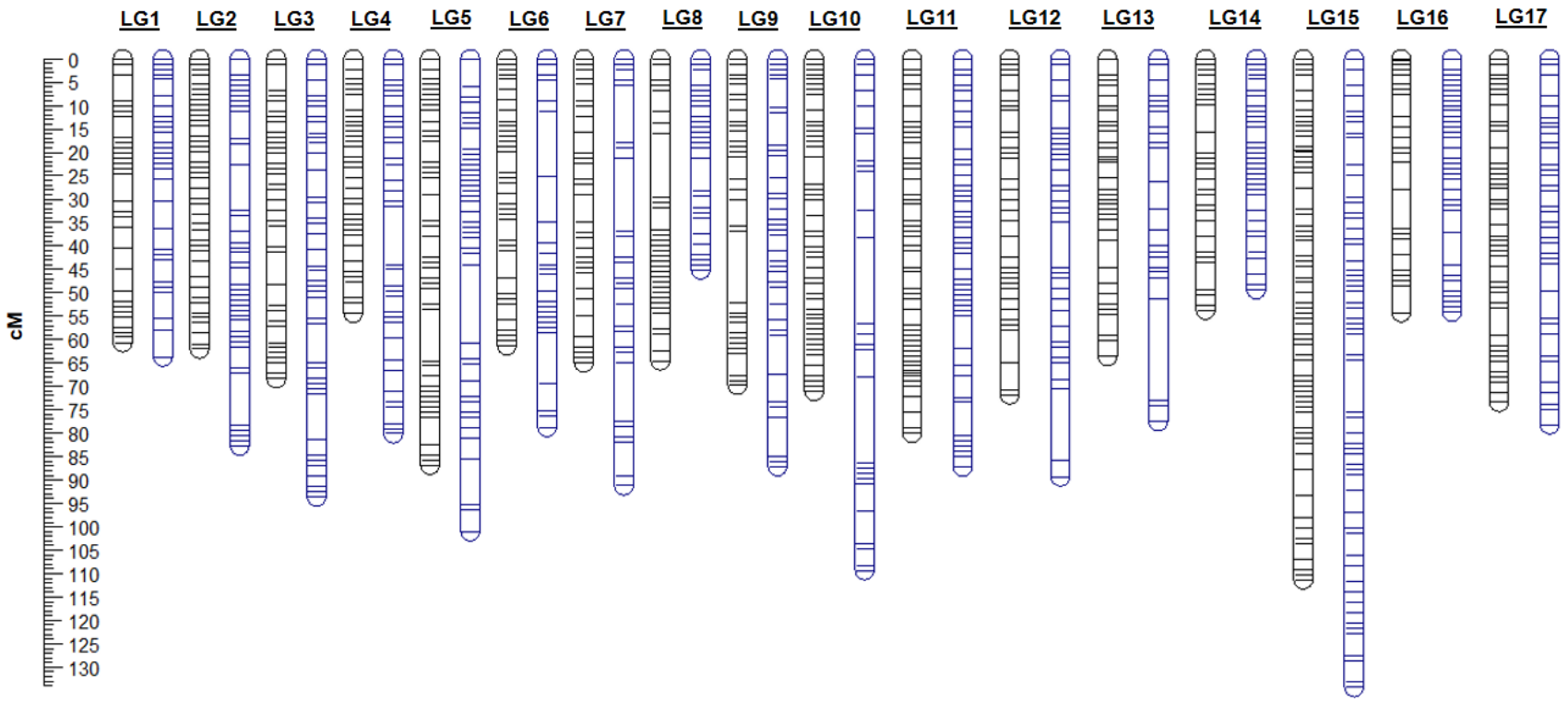


**Figure 4.2. Parental SNP-based genetic linkage maps of 'McIntosh' (in black) and 'M.1' (in blue) across the 17 linkage groups (LGs1-17).** The scale-ruler on left represents map distances in centiMorgan (cM). Each horizontal line inside the LG represent a single locus while the blank white regions inside the LGs indicate gaps between SNP markers.

**Comparison of parental genetic maps to physical positions**

The genetic distances of SNP markers in the parental maps of 'McIntosh' and 'M.1" were plotted against their physical positions on the 'Golden Delicious' double haploid apple genome (GDDH13) v1.1 after which average recombination distances were estimated (Table 4.4; Figure 4.4). Overall, there was collinearity in the SNP marker order between the parental genetic maps and physical positions on the GDDH13 v1.1 reference genome as evidenced by the linearity in the plots. The average recombination rate of the 'McIntosh' and 'M.1' were 0.92 cM/Mb and 1.06 cM/Mb respectively (Table 4.4). The regions on 'McIntosh' are well covered by the markers with LG16 and LG13 having the smallest recombination rate of 0.72 cM/Mb and 0.78 cM/Mb respectively. The regions on 'M.1' were moderately covered with considerable variation in coverage and distribution of SNP markers ranging from LG16 (0.70 cM/Mb) to LG14 (0.77 cM/Mb) (Figure 4.4).

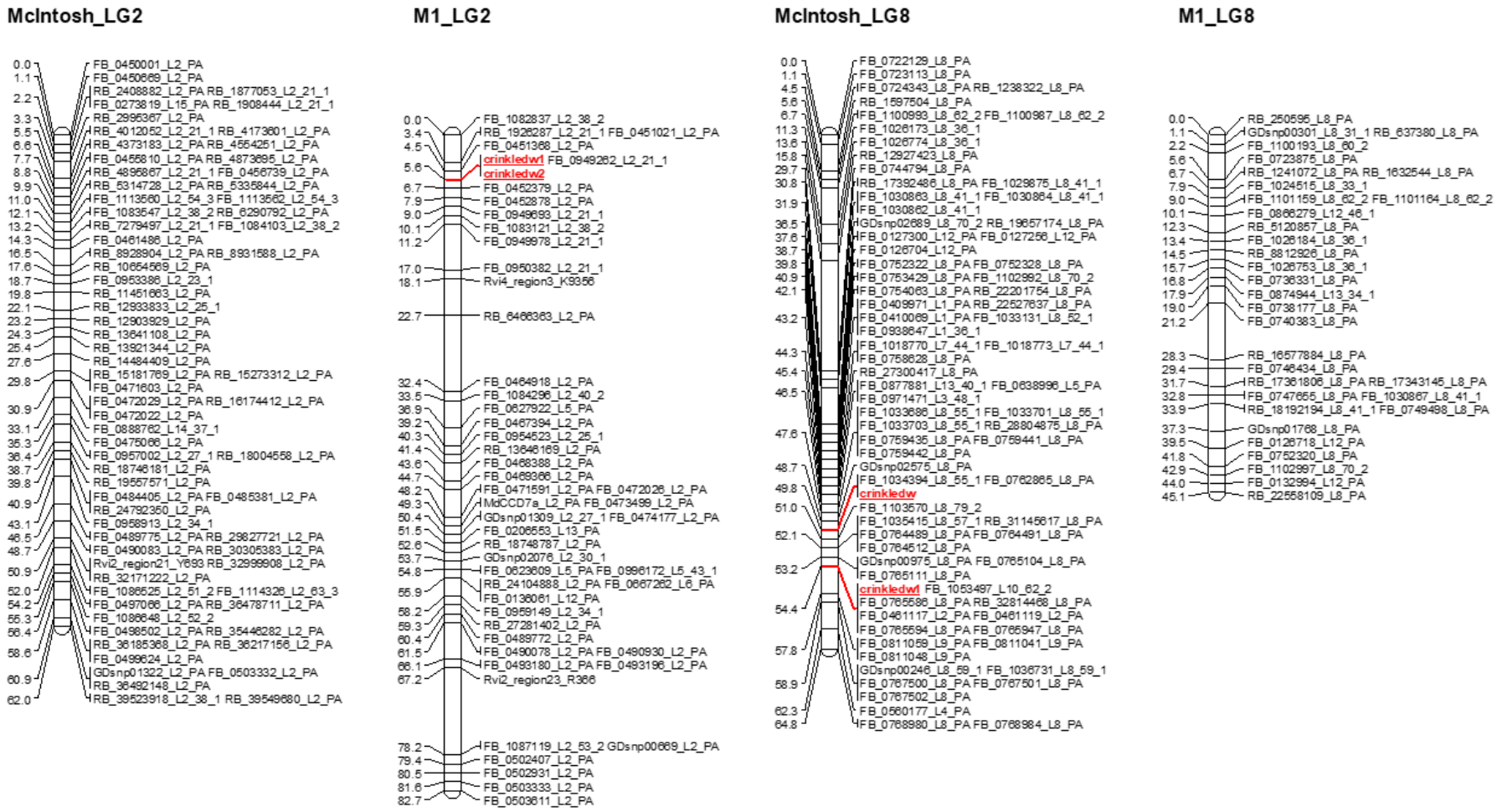


**Figure 4.3. Partial parental genetic maps of 'McIntosh' and 'M.1' outlining linkage groups (LG2 and LG8) with chromosomal regions of mapped crinkle dwarf.** The crinkle dwarf: *crinkledw1* and *crinkledw2* are highlighted in red on LG2 in 'M.1' and on LG8 in 'McIntosh'.

**Table 4.4. Comparison of parental genetic maps to physical positions on the 'Golden Delicious' double haploid v1.1 apple reference genome across the 17 linkage groups.**

| Linkage group | McIntosh parent (♀) | | | M.1 parent (♂) | | |
|---|---|---|---|---|---|---|
| | Linkage map (cM) | [a]Physical map (Mb) | [b]cM/Mb | Linkage map (cM) | [a]Physical map (Mb) | [b]cM/Mb |
| LG 1 | 61.28 | 64.37 | 0.95 | 63.69 | 64.38 | 0.99 |
| LG 2 | 62.66 | 75.09 | 0.83 | 82.69 | 73.43 | 1.13 |
| LG 3 | 67.93 | 74.73 | 0.91 | 93.63 | 74.72 | 1.25 |
| LG 4 | 54.96 | 58.38 | 0.94 | 80.11 | 64.57 | 1.24 |
| LG 5 | 86.67 | 95.86 | 0.90 | 96.42 | 91.61 | 1.05 |
| LG 6 | 62.02 | 74.12 | 0.84 | 78.67 | 73.93 | 1.06 |
| LG 7 | 64.48 | 66.47 | 0.97 | 91.20 | 73.31 | 1.24 |
| LG 8 | 64.46 | 62.43 | 1.03 | 45.13 | 34.80 | 1.30 |
| LG 9 | 68.31 | 71.19 | 0.96 | 87.15 | 75.13 | 1.16 |
| LG 10 | 71.84 | 82.75 | 0.87 | 89.64 | 82.74 | 1.08 |
| LG 11 | 78.65 | 85.44 | 0.92 | 87.16 | 86.03 | 1.01 |
| LG 12 | 71.52 | 65.72 | 1.09 | 85.91 | 65.26 | 1.32 |
| LG 13 | 64.31 | 82.20 | 0.78 | 77.53 | 79.98 | 0.97 |
| LG 14 | 54.33 | 59.22 | 0.92 | 49.33 | 63.71 | 0.77 |
| LG 15 | 109.23 | 109.47 | 1.00 | 96.89 | 109.20 | 0.89 |
| LG 16 | 54.88 | 76.18 | 0.72 | 52.96 | 76.18 | 0.70 |
| LG 17 | 74.24 | 67.79 | 1.10 | 78.18 | 69.15 | 1.13 |
| Total | **1171.75** | **1271.41** | **0.92** | **1336.27** | **1258.13** | **1.06** |

[a]Golden Delicious double haploid (GDDH13) v1.1; [b]Recombination distance/physical distance (recombination rate)

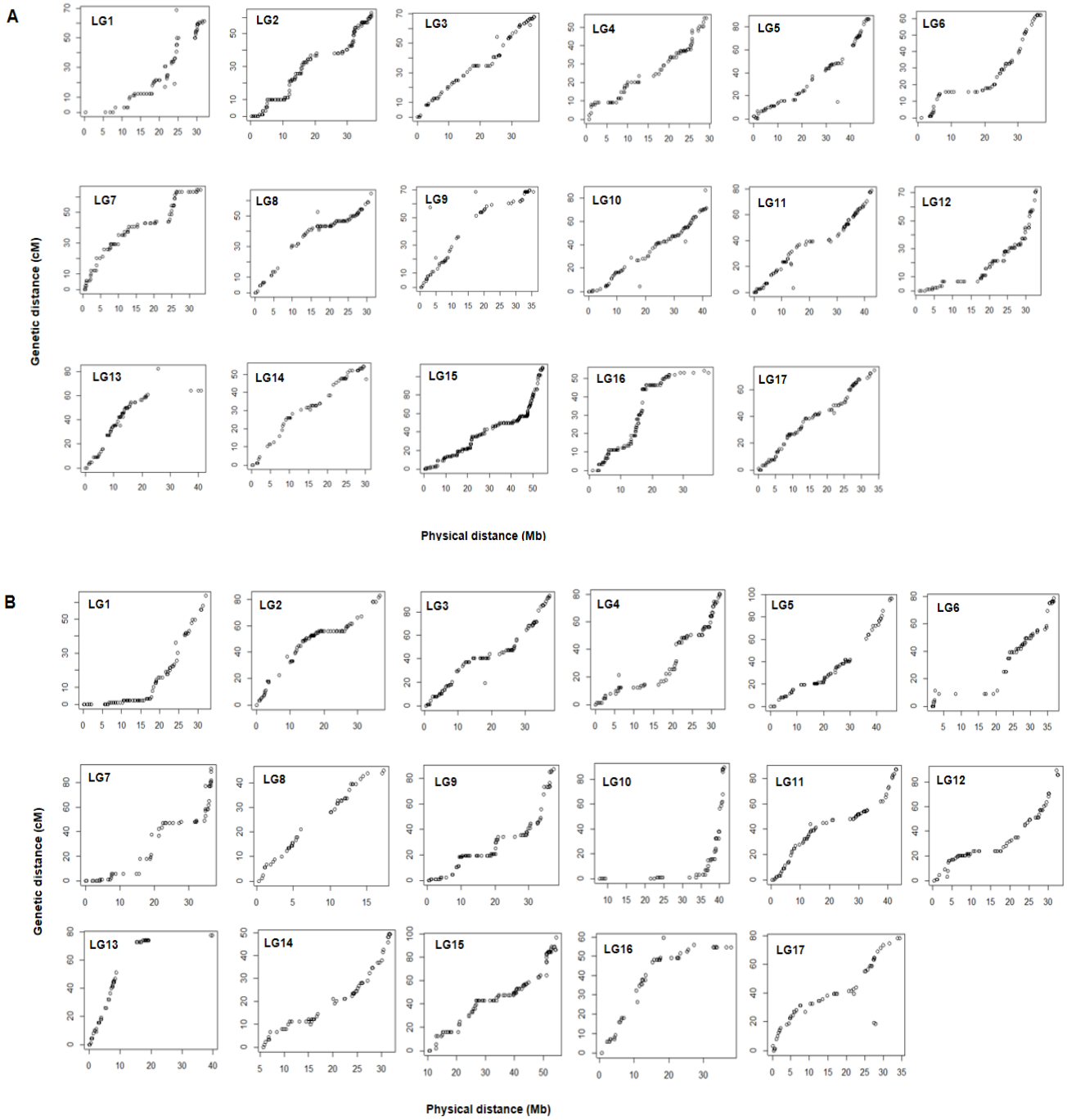


**Figure 4.4. Comparison of the parental genetic maps to physical positions on the 'Golden Delicious' double haploid (GDDH13) v1.1 apple reference genome across the 17 linkage groups.** The sets plots denoted A and B corresponds to the maternal ('McIntosh') and paternal ('M.1') genetic maps. Dots on each plot indicate genetic position of markers in centiMorgans (cM) (left axis), plotted against their estimated physical position in the genome in Megabases (Mb).

**Validation of crinkle dwarf trait using Kruskal-Wallis**

The two highest K* score values of 30.67 ($p < 0.0001$) and 30.44 ($p < 0.0001$) were observed on LG8 of the 'McIntosh' genetic map at genetic position 54.28 cM. The highest K* score of 30.67 ($p < 0.0001$) was at genetic position 54.28 cM and harbored only one marker, SNP_FB_0765586. The sec highest K* was 30.44 ($p < 0.0001$), indicating a significant QTL at 54.28 cM, linked with seven markers, of which six: SNP_FB_1053497, SNP_FB_0461117, SNP_FB_0461119, SNP_FB_0765594, SNP_FB_0765947 and RosBREED_SNP_GP_32814468 were mapped on LG8, of 'McIntosh'. One marker, SNP_FB_1036446 was not mapped. The results obtained from the KW analysis are summarized in Table 4.5 and Figure 4.5.

**Table 4.5. Validation of crinkle dwarf trait on linkage group (LG) 8 in the 'McIntosh' genetic map through Kruskal-Wallis.**

| Trait | [a]QTL | [b]Linkage group | [c]Position (cM) | [d]SNP marker | [e]K* | [f]*p*-value |
|---|---|---|---|---|---|---|
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_0765586 | 30.67 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_1053497 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_1036446 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_0765947 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_0461117 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_0461119 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | SNP_FB_0765594 | 30.44 | 0.0001 |
| *crinkledw1* | crinkledw | LG 8 | 54.28 | RosBREED_SNP_GP_32814468 | 30.44 | 0.0001 |

[a]QTL named using an abbreviation of the trait; [b]Linkage group where QTL was detected; [c]Position of nearest marker to the QTL;
[d]SNP markers closest to the peak of QTL; [e]K* = Kruskal-Wallis statistic score; [f]significance level given by the *p*-value.

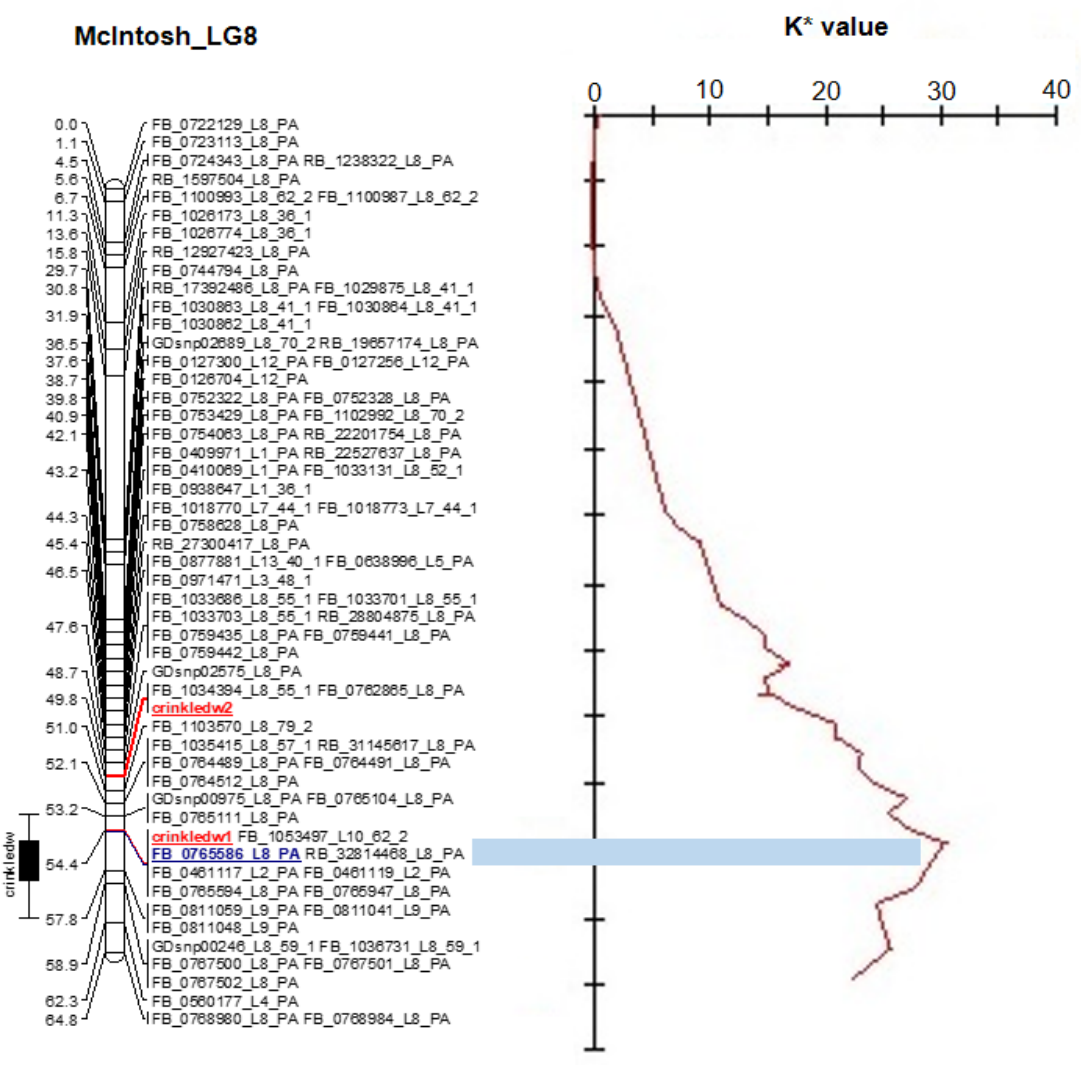


**Figure 4.5. Chromosomal location of crinkle dwarf on the linkage groups (LG) 8 of 'McIntosh' based on Kruskal Wallis non-parametric test.** The names of markers are given on the right of the LG while the map distances in centiMorgan (cM) are indicated on the left. The marker-trait association region is represented by a black solid vertical bar positioned to the left. At the far right, a K* value plot against the genetic distance is indicated by purple traces. The light-blue line corresponds to the highest K* of 30.67 ($p < 0.0001$) with SNP_FB_0765586 being the closest marker.

### 4.4.5 Consensus genetic linkage maps

***Co-segregation analysis based on consensus genetic map and mapping of crinkle dwarf trait***

The construction of the consensus linkage map started with 92 $F_1$ individuals from a cross between 'McIntosh' and 'M.1'. Seven individuals were discarded because of missing data, leaving a total of 85 individuals for the final linkage analysis. The *crinkledw1* and *crinkledw2* loci segregated in a 1:3 (hh:hk) and 3:1 (hk:kk) respectively, and in accord with Mendelian ratio for a monogenic trait segregation. The crinkle dwarf trait co-segregated with the marker GDsnp02575 which segregated 1:1 (lm:ll) together with the other flanking SNP markers (Figure 4.6; Supplementary Table 4.3).

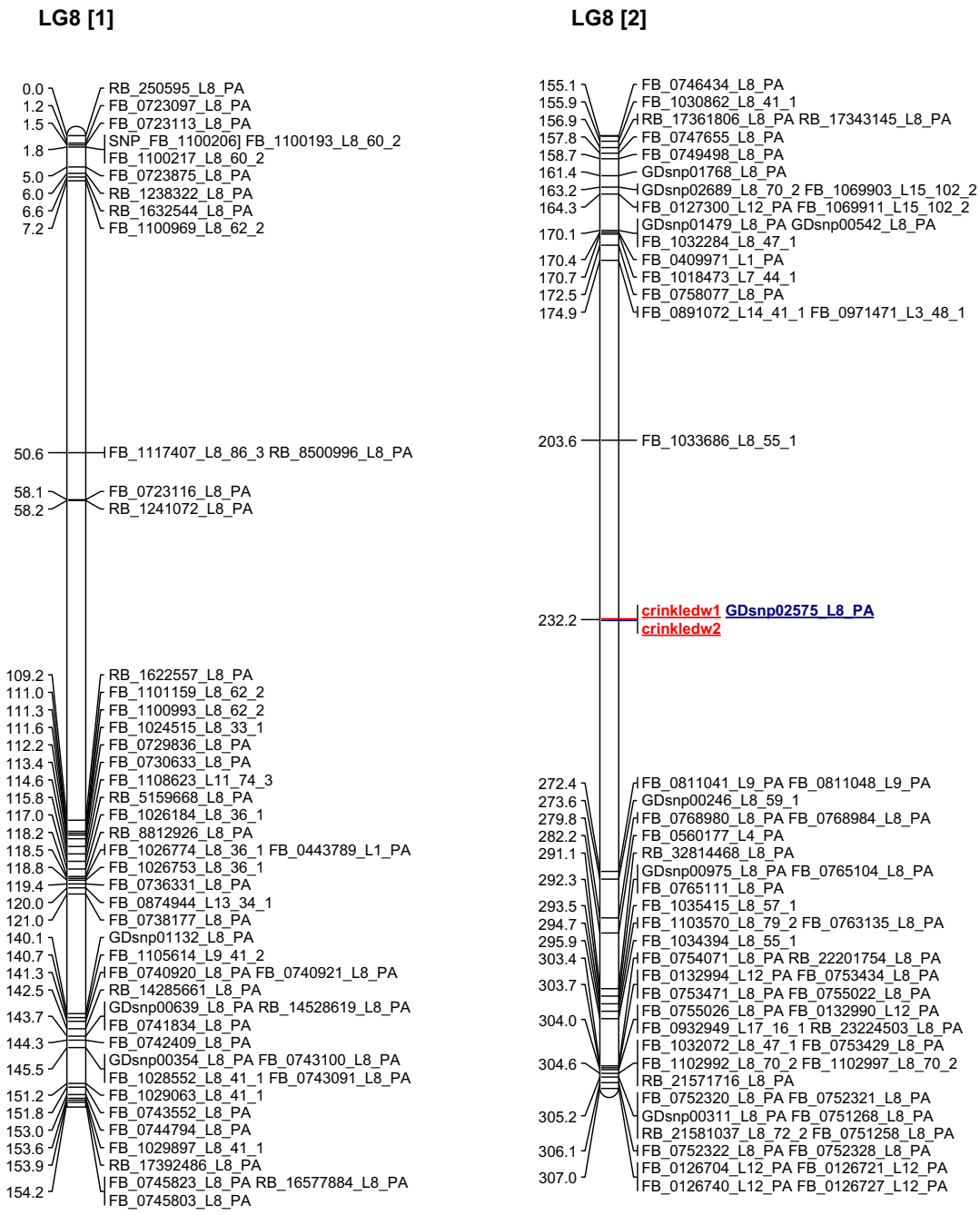


**Figure 4.6 Partial consensus genetic map of 'McIntosh' x 'M.1' on linkage group (LG8) outlining the chromosomal region of mapped crinkle dwarf trait**. The *crinkledw1* and *crinkledw2* are highlighted in red, and the co-segregating marker (GDsnp02575) is highlighted in blue.

Zooming into the two parental maps that comprised of the LG8 integrated maps, it was noted that the crinkle dwarf trait (*crinkledw1* and *crinkledw2*) in 'McIntosh' mapped at 272.801 cM together with GDsnp02575, flanked by markers segregating (ll:lm) at 271.600 cM and 283.228 cM, which is 1.201 cM and 10.427 cM, respectively (Figure 4.6; Supplementary Table 4.4). In 'M.1', crinkle dwarf trait (*crinkledw1* and *cinkledw2*) mapped on position 191.691 cM, and was flanked by the above and below markers at an inter-marker density of 112.272 cM and 69.906 cM, respectively (Supplementary Table 4.5).

A total of 7,950 SNPs were used in the construction of the integrated genetic map of 'McIntosh' x 'M.1' for which 7,875 SNPs were assembled into 17 LGs. On average, the SNP distribution was non-homogeneous in almost all the LGs, with each LG consisting of ambiguous total length of greater than 1000 cM which lead to spurious linkage maps. Therefore, a schematic illustration of the 17 LGs consensus map could not be created. The number of SNP markers per LGs varied and ranged from 329 to 726. The crinkle dwarf trait (*crinkledw1* and *crinkledw2*) mapped to LG8. Linkage group 8 consisted of 472 SNP markers, spanning a total distance of 307.007 cM (Figure 4.6). The greatest inter-marker distance was 43.365 cM. Both *crinkledw1* and *crinkledw2* mapped on 232.246 cM, which co-segregated with the marker GDsnp02575. The inter-marker distance above (e.g. SNP_FB_0759442 at 203.583 cM)) and below (e.g. SNP_FB_0811041 at 272.413 cM) the flanking markers were 28.663 cM and 40.167 cM, respectively. Mapping of crinkle dwarf trait on LG8 was consistent with the results obtained in the parental map of 'McIntosh (Figure 4.3), though the map position were different. However, the crinkle dwarf (*crinkledw1* and *crinkledw2*) traits from the integrated map and that of 'McIntosh' parental map co-segregated with different SNPs (Figure 4.6 and 4.3, respectively). The marker GDsnp02575 in the 'McIntosh' parental map flanked *crinkledw2* at an inter-density distance of 1.11 cM (Figure 4.3).

## 4.5 Discussion

### 4.5.1 Inheritance and co-segregation analysis

Plant growth development are driven by various cellular processes that enable plant organs to develop into various growth architectural traits of different sizes and shapes (Duan *et al.*, 2012; Hollender and Dardick, 2015). Crinkle dwarf seedlings are usually of weak growth, occasionally resulting in seedling death during developmental stages. The crinkle dwarf types may be associated with some lethality, genetic-incompatibility and other modifier genes, and are therefore one of the biggest challenges in apple breeding programmes (Alston, 1976). One of the main objectives in breeding programmes is to elucidate genetic control underlying important traits of interest.

In the current study, the progeny of 'McIntosh' and 'M.1' conformed to the expected 3:1 Mendelian segregation ratio ($\chi^2 = 0.10$, $p = 0.750$, Figure 4.1). Therefore, the crinkle dwarf phenotype is expressed when one of the genes is homozygous recessive *D-ee*, attributed temporarily to *crinkledw* as the dwarfing mutation derived from 'McIntosh'. Our results are consistent with those of Alston (1976), except that 'McIntosh' was not one of the parents and the trait was assigned a gene symbol, *cr*. Hutchinson and Ghose (1937), as cited by (An *et al*., 2015), found a crinkle dwarf in upland cotton, which showed a normal phenotype during the seedling stage and a crinkle dwarf phenotype at the fourth or fifth-leaf stage. The crinkle dwarf mutant was controlled by a completely recessive gene (An *et al.*, 2015). A similar inheritance trend of an undesirable trait was previously studied in apple, for which a virescent trait, characterised by seedlings which exhibited chlorotic leaves at germination, segregated on a 3:1 basis, attributed to the recessive gene (*vir*) (Fernández-Fernández *et al.*, 2014). The former studies indicate that most deleterious trait are controlled by recessive genes (An *et al*., 2015). Therefore, it is suggested in the current study that the crinkle dwarf trait is controlled by only one of the genes, (*D-ee*).

### 4.5.2 Genetic linkage map construction

At present, numerous molecular tools including the NGS high-throughput SNP arrays have been developed and widely used in apple genetic research (Chagné *et al.*, 2012; Bianco *et al.*, 2014, 2016). The primary focus (not limited to) has been in the construction of high-density genetic linkage maps that can be leveraged for the characterisation of genetic resources, marker-trait association and subsequent marker assisted selection, more especially in breeding programmes. In the present study, the apple 20K Infinium® SNP array (Bianco *et al.*, 2014) was applied in the construction of high-density parental genetic linkage maps. The apple SNP arrays are widely been used to discover, explore and manipulate genes ranging from simple and complex traits as well as in genome-wide association studies (Kumar *et al.*, 2013; Falginella *et al.*, 2015; Vanderzande *et al.*, 2017). The utility and power of SNP genotyping relies heavily on marker density and on the ability to assay thousands of SNP markers in parallel (Chagné *et al.*, 2012; Bianco *et al.*, 2014, 2016).

The maternal 'McIntosh' linkage map is comprised of 3190 SNP markers covering a total of 1171.96 cM. The 'McIntosh' genetic linkage map contains only two regions in excess of 10 cM that were devoid of any SNP markers. The largest inter-marker region to which no markers were mapped was 13.66 cM and 15.11 cM observed for LG8 and LG9, respectively. The paternal 'M.1' linkage map comprised 2640 SNP markers spanning a total genetic distance of 1402.44 cM. The 'M.1' genetic linkage map contained nine regions in excess of 10 cM which had no mapped SNP markers, with LG13 being the longest region at 21.67 cM. The genetic linkage maps resulted in good coverage overall, although some of the regions could not be saturated with polymorphic markers in the 'M.1' parental genetic map. The overall coverage of the parental genetic linkage maps constructed in this study was more dense, uniform, and well saturated. These results complement the previous work of Falginella *et al*. (2015), who also employed a similar apple 20K SNP array to construct high-density genetic maps of the parents 'Renetta Grigia di Torriana', ('RGT') and 'Golden Delicious'('GD').

### 4.5.4 Comparison of parental genetic maps to physical position

Overall, the two parental genetic linkage maps of 'McIntosh' and 'M.1' showed a high degree of collinearity against the 'Golden Delicious' double haploid (GDDH13) v1.1 apple reference genome except that a larger inter-marker density of 21.67 cM was observed on LG 13 in 'M.1'. The SNP marker order was in accordance with the GDDH13 v1.1 apple reference genome but with greater marker coverage and density overall.

### 4.5.3 Mapping of crinkle dwarf trait from parental maps

The *crinkledw1 and crinkledw2* mapped on LG8 of the 'McIntosh' genetic linkage map at genetic positions of 54.38 cM and 49.83 cM, respectively. However, in the paternal 'M.1' genetic linkage map, both the *crinkledw1 and crinkledw2* mapped on LG2 at position 5.63 cM. Upon closer inspection, the comparison of genetic length of LG8 of both 'McIntosh' and 'M.1'showed that 'M.1' was shorter. The *crinkledw* trait was mapped at 54.36 cM, a target region which is missing on the LG8 of 'M.1' (Figure 4.3). Therefore, LG8 of the 'M.1' genetic map was compared against the physical map (Figure 4.4) which showed that LG8 was indeed well covered. Poorly saturated genomic regions do not necessarily reflect a shortage of markers, rather a lack of polymorphic markers and the presence of monomorphic SNPs on the target region could therefore explain the former (van Berloo *et al.*, 2008; Antanaviciute *et al.*, 2012). The lack of polymorphic SNP loci may be due to the rootstock genetic background of 'M.1'. The rootstock cultivar Malling 1 ('M.1'), was not included in the development of the apple 20K Infinium SNP array, and has a genetic background more removed from the well characterized scion genotypes included in the development of the SNP array, for which 'McIntosh', scion cultivar was included (Antanaviciute *et al.*, 2012; Chagné *et al.*, 2012; Troggio *et al.*, 2012; Bianco *et al.*, 2014).

Moreover, the missing region in 'M.1' may also be largely due to segregation distortion. In a previous study on cotton, the majority of the regions that showed adverse distortion were mainly skewed towards the male parent (Dai *et al.*, 2017). The occurrence of segregation distortion in at least one locus of a linkage group leads to biased estimates of the distance between loci pairs, decreasing the resolution of the linkage map, and thus complicating the identification and mapping of the trait of interest (Song

*et al.*, 2006). Therefore, one cannot rule out the possibility that some deleterious genes may reside nearby distorted loci. Cheng (1996) also suggested that marker segregation distortion may result from the elimination of certain types of gametes or even of zygotes through lethal factors located in a neighboring region.

Segregation distortion, attributed to lethal genes, has been previously reported in apple (Fernández-Fernández *et al.*, 2014), pear (Montanari *et al.*, 2016) and apricot (*Prunus* spp.) (Vilanova et al., 2003), and in other cereal crops including rice. When a locus is under selection, molecular markers linked to it may also exhibit segregation distortion due to indirect action of the linked loci i.e. lethal-related locus (Song *et al*. 2006). A recent study in pear *(Pyrus* spp.) (Montanari *et al.*, 2016) mapped a hybrid necrosis trait exhibiting similar phenotype to that of the crinkle dwarf studied in this thesis and was mapped on LG2. In the present study, the crinkle dwarf trait on 'M.1' mapped on LG2, even it is assumed to be wrongly placed. However, since apple and pear are homologous, it may be hypothesized that crinkle dwarf trait, a form of hybrid necrosis, may be in fact caused by a lethal gene that acts during prezygotic selection.

Moreover, a similar trend of poor marker coverage of target regions have been reported in similar studies of tomatoes (van Berloo *et al.*, 2008). Nonetheless, the lack of the target region in 'M.1', might have incorrectly placed *crinkledw1* and *crinkledw2* on LG2, however, this needs to be investigated further. In parallel, the progeny of 'McIntosh' x 'M1' is being screened with published apple microsatellites from LG8 and LG2, to fine map the crinkle dwarf trait regions identified here. Concurrently, a number of SNPs that co-segregates with crinkle dwarf trait on LG8 have been selected; and are being validated further on the apple genome. In future, a higher density apple Axiom® 480K SNP array, which include rootstock cultivars upon development, will be applied to increase the resolution of the generated parental maps of 'McIntosh' and 'M.1' using a large number of populations and individuals.

### 4.5.4 Validation of crinkle dwarf trait using Kruskal-Wallis

Taking advantage of the high-density genetic parental maps, *crinkledw1 and crinkledw2* were subjected to QTL analysis using the non-parametric KW test. The KW analysis is well suited for traits with a non-normal distribution due the qualitative nature. In this study, a KW test was performed to confirm the significance of the marker-trait association nearest to the QTL (Kruglyak and Lander, 1995; Kruglyak *et al.*, 1996; Broman, 2002; Fernandes *et al.*, 2007). The results obtained from KW analysis suggests that the position of *crinkledw2* (54.38 cM) may be the correct location for the crinkle dwarf trait (*crinkledw*) which is tightly linked to the SNP marker SNP_FB_0765586. The physical location of the SNP_FB_0765586 is 28,292,148 bp on the GDDH13 v1.1 apple reference genome (Daccord *et al.*, 2017). The co-location of the marker-association validation and *crinkledw* trait is noteworthy, and also showed LG8 has a significant effect on crinkle dwarf growth habit. These results are specific to the 'McIntosh' x 'M.1' mapping population.

Li *et al*. (2014) stated that it is often assumed that a quantitative trait exhibits continuous variation because of the interaction of environmental effects and multiple genes of small and cumulative effects (Li *et al.*, 2014). Wu *et al*. (2014) investigated QTLs of qualitative traits in pear (calyx status, flesh colour, juice content, number of seeds and skin colour) using Kruskal-Wallis, and found that growth-related traits may in part be controlled by a few genes with large effect. The ranges of the K* value, LOD score and PVE obtained in the current study were also comparable to those of Kunihisa *et al*. (2014), who identified marker-trait association for fruit quality traits in Japanese apples, though their study used both normal and non-normally distributed datasets (Kunihisa *et al.*, 2014). The current study is in agreement with other studies in tree species, based on KW, indicating that growth-related traits may in part be controlled by a few genes with large effects (Collard *et al.*, 2005; Kunihisa *et al.*, 2014; Wu *et al.*, 2014)

The signals obtained from KW may indicate the presence of other modifier genes being involved in the crinkle dwarf phenotypic expression. Nevertheless, the limited population size used in the present study could have led to overestimation or underestimation of the probable marker-association even though KW analysis was mainly for validation purposes.

### 4.5.5 Mapping of crinkle dwarf trait from consensus genetic linkage maps

The distribution of markers along linkage groups was not random and there were marker-rich and marker-poor regions evident from the consensus linkage maps (Figure 4.6). Overall, the consensus genetic map showed the highest number of distorted markers, which led to spurious linkage maps. In the present study, based on LG8, the highest inter-marker density was 43.365 cM. The larger gaps obtained from the consensus maps were also reported and are comparable to other studies including maize (*Zea mays*) (Pan *et al.*, 2012) and pigeonpea (*Cajanus cajan*) (Bohra *et al.*, 2012). In an attempt to enhance integration of the consensus map, distorted markers were excluded sequentially, however, this generated even larger gaps. It is important to note that it was for this reason that the mapping strategy utilised in the construction of parental genetic linkage maps excluded most of the SNP marker types <hkxhk> and a few <lmxll> and <nnxnp>.

Segregation distortion was observed in both the parental genetic maps and consensus genetic map with varying degrees of deviation. Segregation distortion is a common phenomenon observed in both intraspecifc and interspecific crosses, however the extent is more in case of interspecific crosses (Bohra *et al.*, 2012). Genetic incompatibilities or hybrid lethal factors resulting in segregation distortion is suggested to be the main evolutionary force driving speciation (Bomblies *et al.*, 2007; McDermott and Noor, 2010). Similar instances of segregation distortion were also reported in apple (Fernández-Fernández *et al.*, 2014) and pear (Montanari *et al.*, 2016). Segregation distortion have also been widely observed in mapping studies in many species including wheat (*Triticum* spp.) (Takumi *et al.*, 2013), rice (*Oryza* spp.) (Harushima *et al.*, 2001, 2002), maize (*Zea mays*) (Lu *et al.*, 2002), cotton (*Gossypium* spp.) (Dai *et al.*, 2017), tomato (*Solanum* spp.). Segregation distortion may result from various factors such as residual heterozygosity, gametophytic or zygotic selections as well as genotyping errors (Song *et al.*, 2006). It could also be due to statistical uncertainty resulting from weak linkages (Alheit *et al.*, 2011). Bodénès (2016), also stated that the accumulation of deleterious mutations is an alternative mechanism that may give rise to segregation distortion, therefore segregation distorted markers may link to genes or traits of interest. Therefore it is advisable to include all the molecular markers in the generation of the genetic linkage maps, as the exclusion of

distorted markers could bias the data and result in the loss of important information (Zamir and Tadmor, 1986; Ouyang *et al.*, 2010; Takumi *et al.*, 2013).

In general, molecular marker coverage and genetic map density are influenced by many criteria such as the type and number of molecular markers, distribution of markers and crossovers in the genome, mapping population size and mapping strategy (Grattapaglia and Sederoff, 1994; Collard *et al.*, 2005). In the present study, 'McIntosh' x 'M.1' consisted of a total of 92 seedling. Therefore, resolving marker order at high-density regions of the genome would require extremely large mapping populations (Alheit *et al.*, 2011). Sample size is crucial for genetic map construction as it affects the power of linkage detection and the accuracy of recombination fraction estimation (Semagn *et al.*, 2006). Random variation and potential biological variation can cause differences in estimated pairwise distances between individual populations, particularly if populations are small in size (Grattapaglia and Sederoff, 1994; Collard and Mackill, 2008; Nadeem *et al.*, 2018). Sample size and genotyping errors are some non-biological factors that can contribute to segregation distortion. Biologically, segregation distortion can be due to selection among gametes and/or zygotes (Alheit *et al.*, 2011; Xu *et al.*, 2013; Reflinur *et al.*, 2014) .

To our knowledge, this is the first study to characterize, map and identify a marker associated with crinkle dwarf phenotype in apple. Therefore, an in-house marker/probe that is specific for crinkle dwarf trait could be designed and used in screening breeding lines and eliminating carriers of crinkle dwarf. The findings of this study provides a platform for future research prospects, though results remains to be confirmed in different apple mapping populations segregating for the crinkle dwarf phenotype.

## 4.6 Conclusion

The parental high-density genetic linkage map for the two parents ('McIntosh' and 'M.1') had sufficient coverage overall, confirming the robustness of SNP markers included in the apple 20K SNP Infinium® array (Bianco *et al.*, 2014). These findings could have practical implications in terms of avoiding raising dwarf seedling associated with crinkled leaves, a deleterious trait, in the near future. Overall, the findings of this study suggests that a lethal-related gene with major effects may control the crinkle dwarf phenotype. The major gene would have a monogenic inheritance pattern with recessive

deleterious allele effects (Charlesworth *et al.*, 1990). A transcriptomic approach is pursued in parallel with the current study to identify differentially expressed genes in normal *versus* dwarf seedlings. Our results imply that caution must be exerted in relation to the segregation distortion observed especially in the 'M.1' paternal genetic map and the consensus genetic map.

## Supplementary Data

Please see back of document

Supplementary Table 4.1 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' genetic linkage map.

Supplementary Table 4.2 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 2 of the 'M.1' genetic linkage map.

Supplementary Table 4.3 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' x 'M.1' consensus genetic linkage map.

Supplementary Table 4.4 Segregation patterns for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' consensus genetic linkage map.

Supplementary Table 4.5 Segregation patterns for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'M.1' consensus genetic linkage map.

## 4.7 References

**Alheit K V., Reif JC, Maurer HP, Hahn V, Weissmann EA, Miedaner T, Würschum T**. 2011. Detection of segregation distortion loci in triticale (x *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. BMC Genomics **12**, 380.

**Alston FH**. 1976. Dwarfing and lethal genes in apple progenies. Euphytica **25**, 505–514.

**An W, Gong W, He S, Pan Z, Sun J, Du X**. 2015. MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum*. BMC Genomics **16**, 886.

**Antanaviciute L, Fernández-Fernández F, Jansen J, *et al***. 2012. Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* infinium whole genome genotyping array. BMC Genomics **13**, 203.

**Atkinson C, Else M**. 2001. Understanding how fruit trees. The Compact Fruit Tree **34**, 46–49.

**Bassil N V., Davis TM, Zhang H, *et al.*** 2015. Development and preliminary evaluation of a 90K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria* × *ananassa*. BMC Genomics **16**, 1–30.

**Bianco L, Cestaro A, Linsmith G, *et al.*** 2016. Development and validation of the Axiom® apple 480K SNP genotyping array. Plant Journal **86**, 62–74.

**Bianco L, Cestaro A, Sargent DJ, *et al.*** 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh.). PLOS ONE **9**, e110377.

**Bohra A, Saxena RK, Gnanesh BN, Saxena K, *et al***. 2012. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theoretical and

Applied Genetics **125**, 1325–1338.

**Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D**. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biology **5**, 1962–1972.

**Bomblies K, Weigel D**. 2007. Hybrid necrosis: Autoimmunity as a potential gene-flow barrier in plant species. Nature Reviews Genetics **8**, 382–393.

**Broman KW**. 2002. QTL mapping in the case of a spike in the phenotype distribution. Genetics **163**, 1169–1175.

**Byrne DH**. 2012. Trends in fruit breeding. In: Badenes, ML and Byrne D, ed. Fruit Breeding. Boston, MA: Springer US, 3–36.

**Chagné D, Crowhurst RN, Troggio M, *et al.*** 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One **7**, e31745.

**Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE, Korban SS**. 2008. Development of a set of SNP markers present in expressed genes of the apple. Genomics **92**, 353–358.

**Chagné D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner SE, Peace CP, Volz RK, Bassil N V.** 2019. Validation of SNP markers for fruit quality and disease resistance loci in apple (*Malus* × *domestica* Borkh.) using the OpenArray® platform. Horticulture Research **6**.

**Charlesworth D, T. Morgan M, Charlesworth B**. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. Evolution **44**, 1469.

**Chen C, Zhiguo E, Lin HX**. 2016. Evolution and molecular control of hybrid incompatibility in

plants. Frontiers in Plant Science **7**, 1–10.

**Cheng R, Saito A, Takano Y, Ukai Y**. 1996. Estimation of the position and effect of a lethal factor locus on a molecular marker linkage map. Theoretical and Applied Genetics **93**, 494–502.

**Clark MD, Schmitz CA, Rosyara UR, Luby JJ, Bradeen JM**. 2014. A consensus 'Honeycrisp' apple (*Malus* × *domestica*) genetic linkage map from three full-sib progeny populations. Tree Genetics and Genomes **10**, 627–639.

**Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK**. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica **142**, 169–196.

**Collard BCY, Mackill DJ**. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **363**, 557–572.

**Costes E, García-Villanueva E**. 2007. Clarifying the effects of dwarfing rootstock on vegetative and reproductive growth during tree development: A study on apple trees. Annals of Botany **100**, 347–357.

**Daccord N, Celton JM, Linsmith G, *et al.*** 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics **49**, 1099–1106.

**Dai B, Guo H, Huang C, Ahmed MM, Lin Z**. 2017. Identification and characterization of segregation distortion loci on cotton chromosome 18. Frontiers in Plant Science **7**, 1–11.

**Duan Y, Li S, Chen Z, *et al.*** 2012. Dwarf and deformed flower1, encoding an F-box protein, is critical for vegetative and floral development in rice (*Oryza sativa* L.). Plant Journal **72**, 829–842.

**Eduardo I, Chietera G, Pirona R, Pacheco I, Troggio M, *et al***. 2013. Genetic dissection of aroma

volatile compounds from the essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. Tree Genetics and Genomes **9**, 189–204.

**Falginella L, Cipriani G, Monte C, Gregori R, Testolin R, Velasco R, Troggio M, Tartarini S**. 2015. A major QTL controlling apple skin russeting maps on the linkage group 12 of 'Renetta Grigia di Torriana'. BMC Plant Biology **15**, 1–13.

**Fallahi E, Colt WM, Fallahi B, Chun I**. 2002. The importance of apple rootstocks on tree growth, yield, fruit quality, leaf nutrition, and photosynthesis with an emphasis on 'Fuji'. HortTechnology **12**, 38–44.

**Fazio G, Wan Y, Kviklys D, Romero L, Adams R, Strickland D, Robinson T**. 2014. *Dw2*, a new dwarfing locus in apple rootstocks and its relationship to induction of early bearing in apple scions. Journal of the American Society for Horticultural Science **139**, 87–98.

**Fernandes E, Pacheco A, Penha-Gonçalves C**. 2007. Mapping of quantitative trait loci using the skew-normal distribution. Journal of Zhejiang University Science B **8**, 792–801.

**Fernández-Fernández F, Padmarasu S, Šurbanovski N, Evans KM, Tobutt KR, Sargent DJ**. 2014. Characterisation of the virescent locus controlling a recessive phenotype in apple rootstocks (*Malus pumila* Mill.). Molecular Breeding **33**, 373–383.

**Font i Forcada C, Guajardo V, Chin-Wo SR, Moreno MÁ**. 2019. Association mapping analysis for fruit quality traits in *Prunus persica* using SNP markers. Frontiers in Plant Science **9**, 2500.

**Foster TM, Celton JM, Chagne D, Stuart Tustin D, Gardiner SE**. 2015. Two quantitative trait loci, *Dw1* and *Dw2*, are primarily responsible for rootstock-induced dwarfing in apple. Horticulture Research **2**, 15001.

**Frett TJ, Reighard GL, Okie WR, Gasic K**. 2014. Mapping quantitative trait loci associated with

blush in peach [*Prunus persica* (L.) Batsch]. Tree Genetics and Genomes **10**, 367–381.

**Ganal MW, Polley A, Graner EM, Plieske J, Wieseke R, Luerssen H, Durstewitz G**. 2012. Large SNP arrays for genotyping in crop plants. Journal of Biosciences **37**, 821–828.

**Grattapaglia D, Sederoff R**. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics **137**, 1121–1137.

**Di Guardo M, Micheletti D, Bianco L, *et al.*** 2015. ASSIsT: An automatic SNP scoring tool for in- and outbreeding species. Bioinformatics **31**, 3873–3874.

**Harrison N, Harrison RJ, Barber-Perez N, *et al***. 2016. A new three-locus model for rootstock-induced dwarfing in apple revealed by genetic mapping of root bark percentage. Journal of Experimental Botany **67**, 1871–1881.

**Hollender CA, Dardick C**. 2015. Molecular basis of angiosperm tree. New Phytologist **206**, 541–556.

**Huq A, Akter S, Sup I, Hoy N, Kim T, Jin Y, Kwon J, Kang K**. 2016. Identification of functional SNPs in genes and their effects on plant phenotypes. Journal of Plant Biotechnology **43**, 1–11.

**Jansen J, De Jong AG, Van Ooijen JW**. 2001. Constructing dense genetic linkage maps. Theoretical and Applied Genetics **102**, 1113–1122.

**Klagges C, Campoy JA, Quero-García J, *et al.*** 2013. Construction and comparative analyses of highly dense linkage maps of two sweet cherry intra-specific progenies of commercial cultivars. PLoS One **8**, e54743.

**Kosambi DD**. 1944. The estimation of map distances from recombination values. Annals of Eugenics **12**, 172–175.

**Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES**. 1996. Parametric and nonparametric linkage

analysis: A unified multipoint approach. American Journal of Human Genetics **58**, 1347–63.

**Kruglyak L, Lander ES**. 1995. A nonparametric approach for mapping quantitative trait loci. Genetics **139**, 1421–1428.

**Kruskal WH, Wallis WA**. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association **47**, 583–621.

**Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C**. 2012. Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). PLoS One **7**, e36674.

**Kumar S, Garrick DJ, Bink MCAM, Whitworth C, Chagné D, Volz RK**. 2013. Novel genomic approaches unravel genetic architecture of complex traits in apple. BMC Genomics **14**, 393.

**Kunihisa M, Moriya S, Abe K, Okada K, Haji T, Hayashi T, Kim H, Nishitani C, Terakami S, Yamamoto T**. 2014. Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. Breeding Science **64**, 240–251.

**Lateef DD**. 2015. DNA marker technologies in plants and applications for crop improvements. Journal of Biosciences and Medicines **3**, 7–18.

**Li X, Singh J, Qin M, Li S, Zhang X, Zhang M, Khan A, Zhang S, Wu J**. 2019. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and GWAS in pear (*Pyrus*). Plant Biotechnology Journal, 0–2.

**Li Y, Wang D, Li Z, Wei J, Jin C, Liu M**. 2014. A molecular genetic linkage map of *Eucommia ulmoides* and quantitative trait loci (QTL) analysis for growth traits. International Journal of Molecular Sciences **15**, 2053–2074.

**Liebhard R, Koller B, Gianfranceschi L, Gessler C**. 2003. Creating a saturated reference map for the

apple (*Malus* x *domestica* Borkh.) genome. Theoretical and Applied Genetics **106**, 1497–1508.

**McDermott SR, Noor MAF**. 2010. The role of meiotic drive in hybrid male sterility. Philosophical Transactions of the Royal Society B: Biological Sciences **365**, 1265–1272.

**Mizuno N, Hosogi N, Park P, Takumi S**. 2010. Hypersensitive response-like reaction is associated with hybrid necrosis in interspecific crosses between tetraploid wheat and *Aegilops tauschii* Coss. PLoS One **5**.

**Montanari S, Bianco L, Allen BJ, *et al.*** 2019. Development of a highly efficient Axiom™ 70 K SNP array for *Pyrus* and evaluation for high-density mapping and germplasm characterization. BMC Genomics **20**, 1–18.

**Montanari S, Brewer L, Lamberts R, Velasco R, Malnoy M, Perchepied L, Guérif P, Durel CE, Gardiner SE, Chagné D**. 2016. Genome mapping of postzygotic hybrid necrosis in an interspecific pear population. Horticulture Research **3**, 15064.

**Montanari S, Saeed M, Knäbel M, *et al.*** 2013. Identification of *Pyrus* single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific *Pyrus* hybrids. PLoS One **8**, e77022.

**Nadeem MA, Nawaz MA, Shahid MQ, *et al.*** 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment **32**, 261–285.

**Van Berloo R, Zhu A, Ursem R, Verbakel H, Gort G, Van Eeuwijk FA**. 2008. Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theoretical and Applied Genetics **117**, 89–101.

**Van Ooijen JW**. 2009. MapQTL®6, Software for the mapping of quantitative trait loci in

experimental populations. Kyazma B.V., Wageningen, Netherlands.

**Van Ooijen JW**. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Research **93**, 343–349.

**Van Ooijen J, Voorrips RE**. 2017. JoinMap5, software for the calculation of genetic linkage maps in experimental populations of diploid species. Kyazma. B.V., Wageningen, Netherlands.

**Orr H, Presgraves D**. 2000. Speciation by postzygotic isolation: forces, genes and molecules. BioEssays **22**, 1085–1094.

**Pan Q, Ali F, Yang X, Li J, Yan J**. 2012. Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS One **7**, 1–9.

**Patocchi A, Frei A, Frey JE, Kellerhals M**. 2009. Towards improvement of marker assisted selection of apple scab resistant cultivars: *Venturia inaequalis* virulence surveys and standardization of molecular marker alleles associated with resistance genes. Molecular Breeding **24**, 337–347.

**Peace C, Bassil N, Main D, *et al.*** 2012. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS One **7**, e48305.

**Peace CP, Bianco L, Troggio M, *et al.*** 2019. Apple whole genome sequences: Recent advances and new prospects. Horticulture Research **6**.

**Di Pierro EA Di, Gianfranceschi L, Guardo M Di, *et al.*** 2016. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Horticulture Research, 16057.

**Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Visser RGF, van de Weg E**. 2014. 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar

widely used in Russia in breeding for biotic and abiotic stresses. Tree Genetics and Genomes **10**, 261–271.

**Pilcher RLR, Celton J, Gardiner SE**. 2008. Genetic markers linked to the dwarfing trait of apple rootstock 'Malling 9'''. Journal of the American Society for Horticultural Science **133**, 100–106.

**Potter D, Eriksson T, Evans RC, *et al.*** 2007. Phylogeny and classification of *Rosaceae*. Plant Systematics and Evolution **266**, 5–43.

**Rafalski A**. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology **5**, 94–100.

**Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z**. 2017. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Molecular Plant **10**, 1047–1064.

**Rebai A**. 1997. Comparison of methods for regression interval mapping in QTL analysis with non-normal traits. Genetical Research **69**, 69–74.

**Reflinur, Kim B, Jang SM, Chu SH, Bordiya Y, Akter MB, Lee J, Chin JH, Koh HJ**. 2014. Analysis of segregation distortion and its relationship to hybrid barriers in rice. Rice **7**.

**Rezvoy C, Charif D, Guéguen L, Marais GAB**. 2007. MareyMap: An R-based tool with graphical interface for estimating recombination rates. Bioinformatics **23**, 2188–2189.

**Semagn K, Bjoernstad A, Ndjiondjop M**. 2006. Principles, requirements and prospects of genetic mapping in plants. African Journal of Biotechnology **5**, 2569–2587.

**Shiratake K, Suzuki M**. 2016. Omics studies of citrus, grape and *Rosaceae* fruit trees. Breeding Science **66**, 122–138.

**Song X-L, Sun X-Z, Zhang T-Z**. 2006. Segregation distortion and its effect on genetic mapping in

plants. Chinese Journal of Agricultural Biotechnology **3**, 163–169.

**Takumi S, Motomura Y, Iehisa JCM, Kobayashi F**. 2013. Segregation distortion caused by weak hybrid necrosis in recombinant inbred lines of common wheat. Genetica **141**, 463–470.

**Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE**. 2012. Apple, from genome to breeding. Tree Genetics and Genomes **8**, 509–529.

**Troggio M, Šurbanovski N, Bianco L, *et al.*** 2013. Evaluation of SNP data from the *Malus* infinium array identifies challenges for genetic analysis of complex genomes of polyploid origin. PLoS One **8**, e67407.

**Vanderzande S, Micheletti D, Troggio M, Davey MW, Keulemans J**. 2017. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. Tree Genetics and Genomes **13**, 125.

**Velasco R, Zharkikh A, Affourtit J, *et al.*** 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genetics **42**, 833–839.

**Verde I, Bassil N, Scalabrin S, *et al.*** 2012. Development and evaluation of a 9K SNP array for peach by internationally coordinated snp detection and validation in breeding germplasm. PLoS One **7**, e35668.

**Voorrips RE**. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity **93**, 77–78.

**Wang J, Li L, Zhang G**. 2016. A high-density SNP genetic linkage map and QTL analysis of growth-related traits in a hybrid family of Oysters (*Crassostrea gigas* × *Crassostrea angulata*) using genotyping-by-sequencing. G3 Genes Genomes Genetics **6**, 1417–1426.

**Webster AD**. 1995. Rootstock and interstock effects on deciduous fruit tree vigour, precocity, and yield productivity. New Zealand Journal of Crop and Horticultural Science **23**, 373–382.

**Wu J, Li LT, Li M, Khan MA, Li XG, Chen H, Yin H, Zhang SL**. 2014. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. Journal of Experimental Botany **65**, 5771–5781.

**Xu X, Li L, Dong X, Jin W, Melchinger AE, Chen S**. 2013. Gametophytic and zygotic selection leads to segregation distortion through in *vivo* induction of a maternal haploid in maize. Journal of Experimental Botany **64**, 1083–1096.

# Chapter 5
# Differential gene expression associated with crinkle dwarf phenotypes in apple based on transcriptome profiling

## 5.1 Abstract

Dwarfism is an important component of plant architecture and significantly affects apple breeding practices and yield. A controlled cross of 'McIntosh' x 'M.1' segregates for normal *versus* crinkle dwarf phenotype. However, crinkle dwarf phenotype, a form of hybrid necrosis, are linked to a significant decrease in quality and yield. It may also be lethal, and thus a highly undesirable agronomic trait. Molecular mechanisms regulating crinkle dwarf phenotype in apple is not well understood. In the present study, transcriptome profiles were generated for pooled tissues (apical buds and young leaves) of normal and crinkle dwarf phenotypes using RNA-Seq high-throughput sequencing. A total of 466,185,872 high-quality clean reads were obtained, and successfully mapped to the apple reference transcriptome, for which 718,945 were properly aligned. A total of 921 (763 up-regulated and 158 down-regulated) differentially expressed genes (DEGs) were obtained. Gene Ontology (GO) enrichment indicated a clear distinction in the cellular component category which were more enriched in the down-regulated transcripts as compared to the up-regulated transcripts. Functional annotation through Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database resulted in successful annotations of 340 (44.56%) up-regulated and 42 (26.58%) down-regulated DEGs which were mapped to 23 and 20 pathways, respectively. These biological pathways demonstrated a complex relationship, which exhibited a high expression of defense signaling and stress-related proteins/enzymes related to crinkle dwarf phenotype. The enzyme lactoperoxidase (KEGG entry EC:1.11.1.7), was the most overly-expressed, in combination with the antioxidant enzyme (glutathione S-transferase, EC:

2.5.1.18), suggesting an involvement of reactive oxygen species (ROS) in this process. A receptor-like kinase (RLK), protein serine/threonine phosphatase (EC:3.1.3.16) was notably upregulated. There was also an up-regulation of pathogenesis-related proteins such as ChiC (EC:3.2.1.14) and pectin demethoxylase (EC:3.1.1.11), as well as up-regulation of alpha-linolenic acid, a precursor of the phytohormone jasmic acid. On the contrary, suberin-related feruloyl transferase, fatty acids, flavonoid and diterpenoid (2-beta dioxygenase, GA2ox) biosynthesis were down-regulated. Collectively, these results suggest that the mechanisms responsible for crinkle dwarf phenotype are similar to those activated in response to pathogenic attack. Therefore, peroxidase together with the serine/threonine phosphatase catalytic subunit (PP2) may be possible causal enzymes/proteins related to the molecular mechanism underlying crinkle dwarf phenotype. This may imply that an autoimmune response might have been triggered by the allele incompatibilities, in this case between 'McIntosh' and 'M.1'.

## 5.2 Introduction

During evolution, ancestral species may diverge into several species that become genetically isolated from one another and develop a reduced capacity for hybridization, resulting in hybrid incompatibility (Orr, 1996; Rieseberg and Willis, 2007; Rieseberg and Blackman, 2010). Hybrid necrosis, a form of hybrid incompatibility, as explained by the Dobzhansky-Muller model is attributed to the theoretical explanation for postzygotic isolation, arising from deleterious interactions between genes at different loci (Orr, 1996; Orr and Presgraves, 2000; Bomblies *et al.*, 2007). Hybrid necrosis (hybrid weakness and/or hybrid lethality) is often associated with cell death, tissue necrosis, wilting, yellowing, chlorosis, dwarfism and reduced growth rate, and in some cases lethality (Burke and Arnold, 2001; Bomblies and Weigel, 2007). Hybrid necrosis is generally triggered by an autoimmune response, and the causal genes encode disease resistance-related proteins in higher plants such as *Arabidopsis* and lettuce (*Lactuca* spp.) (Bomblies *et al.*, 2007; Alcázar *et al.*, 2009, 2010; Jeuken *et al.*, 2009). Moreover, studies on tomatoes (*Solanum lycopersicum*) and *Arabidopsis* also indicated that some forms of hybrid necrosis result from incompatible alleles encoding pathogenesis-resistance (PR) proteins inducing an autoimmunity-like response when in combination with other genes (Krüger *et al.*, 2002; Bomblies *et al.*, 2007; Alcázar *et al.*, 2009)

Dwarfism has been the subject of intensive research in the last decade, and a major contributor to the 'green revolution' (Khush, 2001; Hedden, 2003; Elias *et al.*, 2012). However, dwarfism cannot always be linked to increased yield. Plants sacrifice their growth for different defense responses (Huot *et al.*, 2014). As plants are sessile, they have evolved a highly effective multilayered innate immune system to combat pathogen attack (Eichmann and Schäfer, 2015).

Plant innate immunity employs a two-interconnected system which are activated by cell surface or intracellular receptors, pathogen-associated molecular pattern (PAMPs)-triggered immunity (PTI) and an effector-triggered immunity (ETI) pathways (Jones and Dangl, 2006; Dodds and Rathjen, 2010; Li *et al.*, 2016). The PTI is the first line of defense, which requires membrane receptor proteins known as pattern recognition receptors stimulated by chitin, flagellin or elicitors (Thomma *et al.*, 2011). On the contrary, ETI is the sec line of defense that requires intracellular receptors of pathogen virulence molecules called effectors, and are triggered by pathogenesis-related (PR) proteins (Jones and Dangl, 2006; Macho and Zipfel, 2014). Moreover, ETI involves the activation of further resistance genes often accompanied by the hypersensitive response (HR) (a form of programmed cell death that effectively restricts pathogen growth) and systemic acquired resistance (SAR) (Hammerschmidt, 1999; Gruner *et al.*, 2013; Gao *et al.*, 2015). Thomma *et al*. (2011) also stated that, an overlapping set of downstream immune responses results from the PTI/ETI continuum. This includes the activation of multiple signaling pathways involving mitogen-activated protein kinases (MAPK), reactive oxygen species (ROS), defense hormones (such as salicylic acid, jasmonic acid and ethylene) and the production of antimicrobial compounds (O'Brien *et al.*, 2012; Macho and Zipfel, 2014; Zhang *et al.*, 2016).

In many plant species, the early stages of pathogen infection result in a phenomena called oxidative burst (Bindschedler *et al.*, 2006; O'Brien *et al.*, 2012). This involves the induction of genes or proteins that promote the accumulation of reaction oxygen species (ROS). Reactive oxygen species are a product of normal cellular metabolism that regulate plant growth and development (Mittler *et al.*, 2004; Foyer and Noctor, 2009; Kärkönen and Kuchitsu, 2015). Reaction oxygen species are considered as unavoidable by-products of aerobic metabolism whose production is confined to cellular compartments with strong electron flow such as chloroplast, mitochondria and peroxisomes (Mittler *et al.*, 2004; Choudhury *et al.*, 2013). However, under stress conditions, the balance between ROS production and

elimination is disturbed in cellular components of plants (Karuppanapandian *et al.*, 2011; Kärkönen and Kuchitsu, 2015). Reactive oxygen species include the superoxide radical anion ($O_2^{\bullet-}$), hydroperoxyl radical ($HO_2$), hydroxyl radical ($HO^{\bullet}$), hydrogen peroxide ($H_2O_2$) and singlet oxygen ($^1O_2$), and all are cytotoxic to plants (Mittler, 2002; Scandalios, 2005; Jones and Dangl, 2006; Karuppanapandian *et al.*, 2011). Amongst the ROS, hydrogen peroxide ($H_2O_2$) is fairly stable, and is considered the predominant ROS involved in cellular signaling (Kuźniak and Urbanek, 2000; Karuppanapandian *et al.*, 2011). Therefore, the accumulation of ROS within infected cells is a central component of plant defense signaling pathways which ultimately promotes apoptosis (cell death) (Scandalios, 2005; Jones and Dangl, 2006; Kaurilind *et al.*, 2015). In turn, plants must also scavenge ROS to prevent cell damage by synthesizing antioxidants (such as ascorbate peroxidase, glutathione peroxidase, superoxide dismutases, catalases and peroxidases, amongst others (Scandalios, 2005; Bindschedler *et al.*, 2006).

Apple (*Malus pumila* Mill.) is one of the most important cultivated tree fruit crops worldwide (Fernández-Fernández *et al.*, 2014). Modern practices for controlling growth and vigour in apple breeding programs rely primarily on the use of dwarfing rootstocks (Atkinson and Else, 2001; Costes and García-Villanueva, 2007). The control of tree size is critical for the optimization of fruit productivity and efficiency of orchard breeding strategies (Webster, 1995; Fallahi *et al.*, 2002; Byrne, 2012; Hollender and Dardick, 2015). Three rare forms of dwarf types (early, crinkle and sturdy dwarfs) have been described in apple (Alston, 1976). Crinkle dwarfs are characterised by dwarf seedlings associated with small-rounded, dark-green crinkled leaves, recognisable at four to six weeks after germination while more distinct from twelve weeks (Alston, 1976).

Currently, little is known about the molecular characteristics associated with the crinkle dwarf phenotype in apple. More recently, the availability of the apple reference genome and transcriptome, together with the rapid developments in high-throughput approaches and bioinformatics analysis, has enabled in-depth transcriptome profiling of these phenotypes (Wang *et al.*, 2009; Velasco *et al.*, 2010; Bai *et al.*, 2014; Conesa *et al.*, 2016; Daccord *et al.*, 2017). RNA-Seq is now regularly applied in the identification of genes that are differentially expressed between two or more biological conditions (Wang *et al.*, 2009; Costa-Silva *et al.*, 2017). Several studies in apple (*Malus* spp.) (Krost *et al.*, 2012),

pear (*Pyrus* spp) (Bai *et al.*, 2013; Ou *et al.*, 2015) and apricot (*Prunus* spp) (Zhang *et al.*, 2018) have demonstrated the use of RNA-Seq to successfully elucidate key molecular changes in gene expression of apical buds and young leaves.

The aim of this study was to generate transcriptome profiling of six pools (3 normal and 3 crinkle dwarf) consisting of meristematic tissues (apical buds and young leaves), collected at different developmental stages, using RNA sequencing technology. This could assist in elucidating the molecular mechanism underlying crinkle dwarf phenotype by identifying differential expressed genes between the contrasting phenotypes. Additionally, functional enrichment analysis was pursued in order to identify potential candidate genes by studying biological pathways.

## 5.3 Materials and Methods

### 5.3.1 Plant Material

The study population consisted of 118 seedlings of the cross 'McIntosh' (♀) × 'M.1' (♂) that segregates for the crinkle dwarf phenotype. This population was raised in 2016 and grown on their own roots under natural photoperiodic conditions and housed at the glasshouse at the Agricultural Research Council (ARC) Infruitec-Nietvoobij' Bien Donné Research Farm, Groot Drakenstein, Western Cape, South Africa [(33°83'33"32.06 (S); 18°98'33"33.59 (E)]. The population was visually scored throughout the development stage and classified qualitatively as either "normal", having standard height and healthy leaves, or "crinkle dwarf", with seedlings of short stature with abnormal dark green crinkled leaves usually bristle in texture.

### 5.3.2 Experiment Rationale and Design

Apple trees undergo suspension of vegetative growth during winter seasons, called dormancy (Legave *et al.*, 2015). Lang *et al*. (1987) defined "dormancy" as a temporary suspension of visible growth of any plant structure containing a meristem. To overcome dormancy, plants have to satisfy their chilling requirements to initiate spring bud break (Labuschagné *et al.*, 2002; van Dyk *et al.*, 2010; Atkinson *et al.*, 2013). However, a marked increase in temperatures during winter seasons has been observed in the Western Cape Province, a major apple growing region, which could be related to the effect of global

warming (Cook and Jacobs, 2000; Labuschagné et al., 2002; van Dyk et al., 2010). Therefore, insufficient winter chilling may alter growth synchronization patterns especially in the initiation development of plant organs as a result of irregular and suboptimal budbreak, with negative impacts on fruit crop yields (Atkinson et al., 2013; Porto et al., 2015). For this reason, the progeny of 'McIntosh' x 'M.1' were placed in cold storage to compensate for the lack of sufficient natural chilling requirements prior to experimental initiation.

In the winter of 2017, 118 seedlings were placed in cold storage (0 – 4 °C) for about 45 days (seven weeks), a time equivalent to about 1100 hr (1100 chilling units, CU). During this time, the seedlings were watered once a week. The amount of chilling units (CU) was calculated according to the Utah Chill Unit model developed by Richardson *et al*. (1974) as cited by Labuschagné *et al*. (2002) and Melke and Fetene, (2014). Briefly, an hour at a temperature between 1.5 °C and 2.4 °C contributes to 0.5 CU and equates to 1 hour × 0.5 of chill unit requirement. On the other hand, an hour at temperatures between 2.5 ºC and 9.1 ºC, provides 1.0 CU, which equates to 1 hour x 1 CU (Melke and Fetene, 2014). At the end of artificial winter chilling, ie. days after cold storage, on the 12$^{th}$ September 2017 (spring season), the progeny seedlings were transferred back to the glasshouse and allowed to grow under natural conditions.

***Collection of tissue samples***

To create inventories for transcriptome profiling, four developmental stages containing vegetative meristematic tissues were collected, and are represented in Figure 5.1: Stage I (silver tip), showing silver to light-greenish tip surrounded by bud scales; stage II (green tip), the apical growth point had grown longer and wider, showing distinct green tip; stage III (semi-folded leaves), showing close to slightly folded leaves, which were enlarged, and Stage IV (young leaves), clearly visible unfolded leaves.

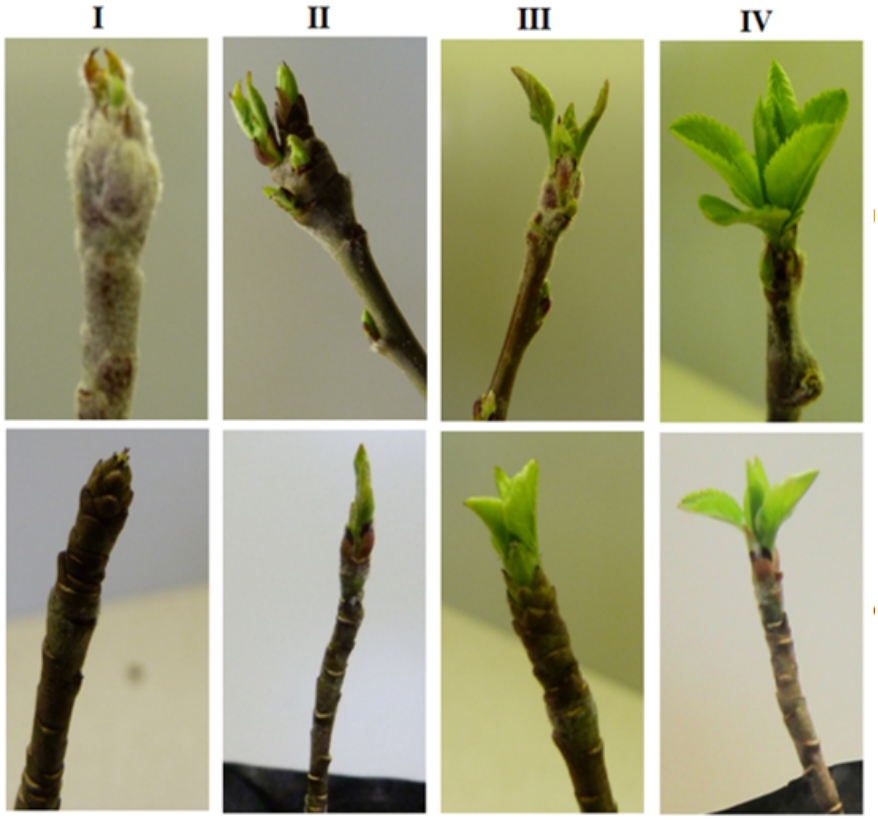


**Figure 5.1 Meristematic tissues at four developmental stages between normal (top row) and crinkle (bottom row) phenotypes.** Stages (I-IV) comprised: I (silver tip), silver-greenish tip as the first visible swelling of the bud showing pinking-silver to green tip; II (green tip), buds gradually growing in length and diameter showing distinct green tip; III (semi-folded leaves), first true leaves separating; IV (young leaves) young unfolded leaves.

The sampling took place in Spring, between 15$^{th}$ September – 23$^{rd}$ September 2017. All tissue materials were sampled by one person, each day at the same time (09:00 – 10:00 AM). To minimize variation, all tissue samples obtained from the same developmental stages were collected on the same day. All tissues per developmental stage were randomly sampled from an independent seedling, in triplicate (representing biological replicates), and immediately frozen in liquid nitrogen and stored at –80 °C. Overall, 24 seedlings were sampled (12 normal and 12 crinkle dwarf phenotypes).

### 5.3.3 RNA extraction and quality assessment

Total RNA was extracted from 100 mg finely ground leaf tissues using the PureLink® Plant RNA Reagent (Catalog Number 12322-012) (Life Technologies, Austin, USA), following the small scale RNA extraction protocol according to manufacturer's instruction. Total RNA was suspended in nuclease-free water and quantified using a BioDrop spectrophotometer (BioDrop Technology, Rockland, UK), following the manufacturer's instructions. The quality of total RNA was estimated by the ratios of absorbance at 260 to 280 nm and 260 to 230 nm. The quality and integrity were assessed using the RNA Nano 6000 Assay Kit and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) at the Central Analytical Facilities (CAF) of Stellenbosch University. The total RNA was shipped on dry ice to the Agricultural Research Council Biotechnology Platform, Pretoria, South Africa, for cDNA library construction.

***Pooling strategy***

Each of the six (three normal, three crinkle dwarf) pooled samples consisted of an equimolar concentration (250 ng) of total RNA extracted from all four developmental stages per replicate, labelled "pool1 to pool6". For example, (Pool1 = stages I, II, III, IV, replicate 1; Pool2 = stages I, II, III, IV, replicate 2; Pool3 = stages I, II, III, IV, replicate 3) for normal phenotypes. Similarly, in crinkle dwarf phenotype, (Pool4 = stages I, II, III, IV, replicate 1, Pool 5 = stages I, II, III, IV, replicate 2 and Pool6= stages I, II, III, IV, replicate 3).

### 5.3.4 cDNA Library Construction and Illumina Sequencing

Prior to cDNA library construction, the total RNA was pooled as per the pooling strategy described above. The cDNA libraries were prepared using the TruSeq RNA sample Prep Kit v2 (Illumina, San Diego, California, USA) according to the manufacturer's instructions. The libraries were quality checked and quantified using the Agilent BioAnalyzer 2100 system (Agilent Technologies, Santa Clara, California, USA). The cDNA libraries were freeze-dried and subsequently shipped to Fondazione Edmund Mach (FEM), San Michelle, Italy, for sequencing. The six cDNA libraries were sequenced twice in a single lane on an Illumina HiSeq 2500 sequencer, generating 101 bp paired-end reads.

### 5.3.5 Quality Assessment, Filtering and Alignment

The paired-end raw reads (FASTQ sequence files) were initially checked for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw reads in Fastq format were subsequently trimmed and filtered using Trimmomatic version 0.38 (Bolger *et al.*, 2014) in paired-end mode with phred +64 quality scores. The ILLUMINACLIP trimming was performed using the parameters; LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:70 (Bolger *et al.*, 2014). Quality of the trimmed reads was assessed through FastQC, ensuring suitable quality for downstream analysis. Bowtie2 mapping aligner (Langmead, 2013) was used to build the index and align reads to the apple reference transcriptome, (GCF_000148765.1_MalDomGD1.0_genomic.gff.gz) (Bai *et al.*, 2014).

After the mapping procedure, a Python package HTSeq (Anders *et al.*, 2015) was used to count unique fragments mapping in each genomic feature using the intersection-nonempty mode. The transcript abundances were represented as fragments per kilobase of transcript sequence per million base pairs (FPKM), i.e. values normalised by transcript length and total number of reads per sample (Mortazavi *et al.*, 2008; Trapnell *et al.*, 2010).

### 5.3.6 Differential Expression Analysis and Functional Enrichment

The R programming environment (v3.6.1) (R Core Team, 2018) and the Bioconductor packages (Anders and Huber, 2010) were used to process raw counts and perform differential expression analysis. More specifically, differential expression of genes were assessed using DESeq2 (v1.24.0) (Love *et al.*, 2014). The p-values were adjusted through false recovery rate (FDR) following the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). The corrected $p$-value < 0.05 and the log2fold change of 1.5 (> +1.5 and < -1.5) were set as the threshold to determine significant differential expression. Positive fold changes indicate up-regulation and negative values indicate down-regulation.

Gene Ontology (GO), Clusters of Orthologous Group (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using OmicsBox-Blast2Go (Trial *v*)

(www.biobam.com/omicsbox) (Tatusov *et al.*, 2000; Conesa *et al.*, 2005, 2016; Conesa and Götz, 2008; Götz *et al.*, 2011; Kanehisa *et al.*, 2016; Huerta-Cepas *et al.*, 2019). Briefly, nucleotide sequences for all differential expressed genes were obtained from NCBI-BatchEntrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez), and subsequently imported into Blast2GO (Conesa *et al.*, 2005; Götz *et al.*, 2008). The function BLASTX was performed against the nonredundant (nr) database for all sequences using an ExpectValue cutoff of $1.0E^{-3}$ and a HSP length cutoff of 33. The next step was GO-Mapping performed using an E-Value-Hit-Filter of $1.0E^{-6}$ with annotation cutoff of 55, GO Weight of 5, and an Hsp-Hit coverage cutoff of 0. The downstream analysis was GO annotations according to the three main terms of GOs (biological process, molecular function and cellular component), an Interpro scan and Go-Slim using default parameters. Finally, enzyme mapping of annotated sequences was analysed on the Kyoto Encyclopedia of Genes and Genomes (KEGG), through add-in on OmicsBox-Blast2GO in defining the main metabolic pathways (Kanehisa and Goto, 2000; Conesa and Götz, 2008; Kanehisa *et al.*, 2016).

## 5.4 Results

### 5.4.1 RNA extraction and cDNA library construction

All extracted total RNA samples were of high quality, with A260/280 values ranging between 1.88 and 2.2, and A260/230 values ranged from 2.06 to 2.18, indicating that RNA was of high purity and without contamination by organic solvents and secary metabolites. The RIN scores ranging from 7.6 to 9.9, generated after analysis with the Agilent 2100 Bioanalyzer were also indicative of high quality RNA. The subsequent cDNA libraries constructed from the six pools resulted in decent concentrations of between 40.6 and 57.0 ng/µL. Only Pool1's concentration was relatively low at 22.0 ng/µL (Supplementary Table 5.1.). Overall, the total RNA and cDNA libraries were considered to be of good quality and suitable for downstream sequencing.

### 5.4.2 Sequencing, Filtering, and Alignment

The paired-end (PE) sequencing of the six pooled cDNA libraries, resulted in a total of 654,173,828 raw reads (i.e. 41.14 GB of raw data). Post-quality filtering, 466,185,872 (71.26 %) high-quality reads

were obtained, with GC content of 48 %. The clean reads were mapped to the publically available apple reference transcriptome at an average of 87.80 % and resulted in 718,945 properly paired transcripts (Table 5.1; Supplementary Table 5.2). Overall, the sequenced reads between the normal and crinkle dwarf were similar in terms of quality and coverage and, therefore, suitable for downstream analysis.

**Table 5.1 Summary of raw and mapped sequenced reads obtained from pooled samples of normal and crinkle dwarf phenotypes.**

| Sample Description | Normal | Crinkle | Total |
|---|---|---|---|
| Total raw reads | 346 037 042 | 308 136 786 | 654 173 828 |
| Total clean reads | 245 394 656 | 220 791 216 | 466 185 872 |
| Total mapped reads | 215 541 933 | 193 769 269 | 409 311 202 |
| Total reads properly aligned | 187 455 038 | 168 840 530 | 356 295 568 |
| Properly paired transcripts | 356 330 | 362 615 | 718 945 |
| GC content (%) | 48 | 48 | |
| Average mapping (%) | 87.81 | 87.78 | |

### 5.4.3 Differential Expressed Genes (DEGs)

Properly paired mapped reads

Initially, 72,569 transcripts (raw counts) were regarded as differentially expressed between normal and crinkle dwarf phenotypes. After data normalisation, the transcripts were filtered to 60,323. The total number of DEGs after correction (p-adjusted < 0.05) was 6,729 with 4,011 and 2,718 being up-regulated and down-regulated, respectively. A selection criterion of p-adjusted < 0.05, log2foldchange > +1.5 and < -1.5, FDR < 0.01 was chosen in selecting significantly differentially expressed transcripts. A total of 1,898 (1,740 up- and 158 down-regulated) DEGs were obtained, of which the distribution is represented in Figure 5.2. From the 1,740 up-regulated DEGs, only 763 matched the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Therefore, 763 up-regulated and 158 down-regulated DEGs (Figure 5.3) were used in further analysis.

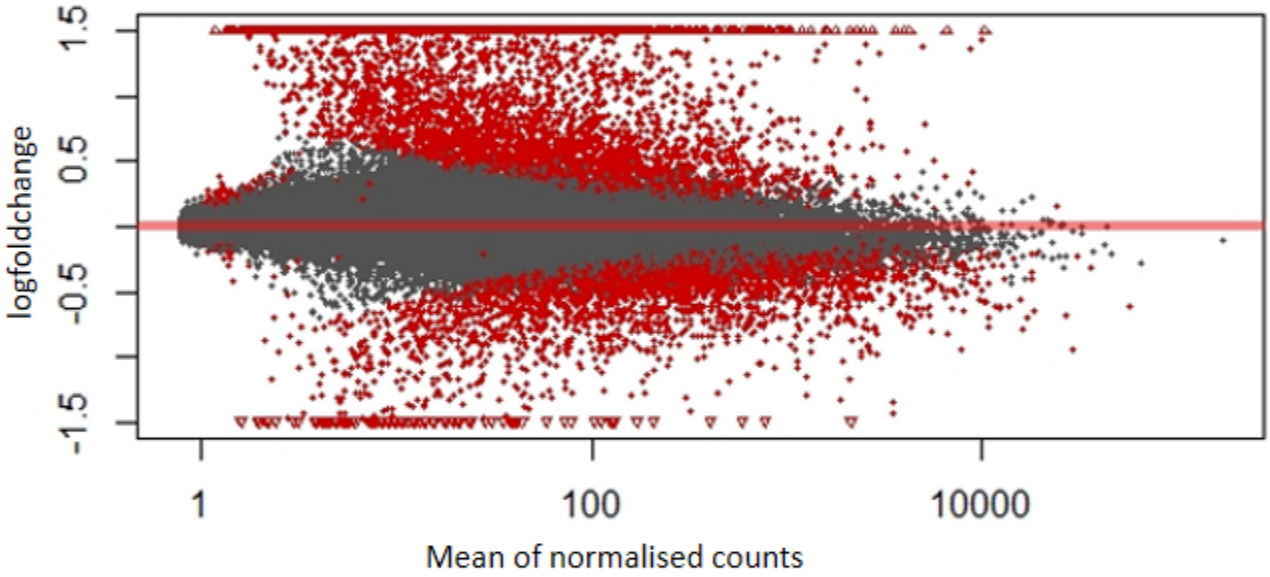


**Figure 5.2 MA plot of differentially expressed genes identified between the normal and crinkle dwarf phenotypes.** Data points represent individual transcript responses plotted as logfoldchange (log2foldchange) against the mean of normalised counts with a negative change representing the down-regulated genes and a positive change representing the up-regulated genes. Grey and red points represent transcripts having False Discovery Rate (FDR) > 0.01 and < 0.01 respectively.

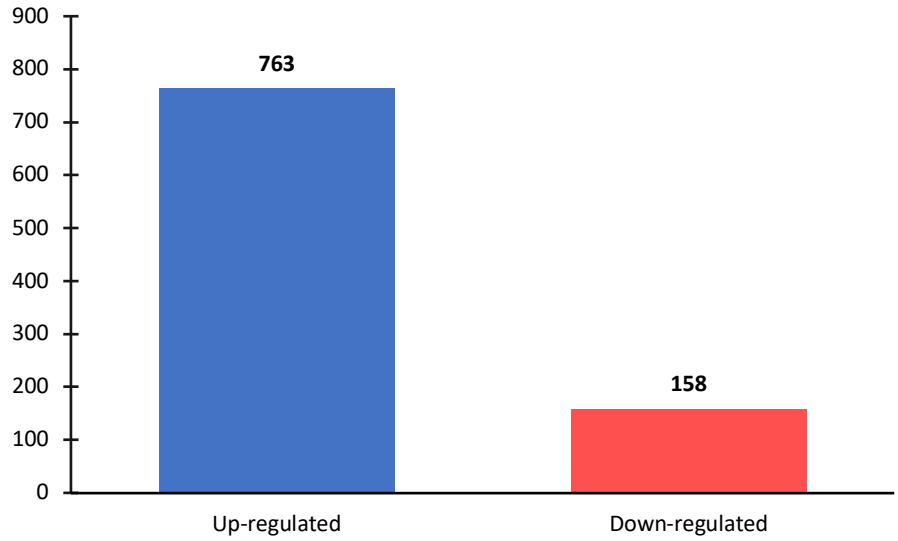


**Figure 5.3 Number of up- and down-regulated differential expressed genes (DEGs) between normal and crinkle dwarf phenotypes.** The DEGs are significant at p-adjusted < 0.05, log2foldchange > 1.5 and < -1.5, FDR < 0.01.

### 5.4.4 Functional Enrichment Analysis

***Gene Ontology (GO) Enrichment***

Gene ontology covers three domains: biological process, molecular function and cellular category. The GO enrichment analysis was performed separately on the 763 up-regulated and 158 down-regulated transcripts' respective datasets using the GO classification system. The results presented herein represent GO terms analysed at level 2. The biological process and molecular function categories showed similar GO terms distribution between the up- and down-regulated transcripts. One, notable sub-category/ difference was the 'response to stimulus", which is a defense-related category. The top two sub-categories under biological processes, were "metabolic process" and "cellular process". The top two sub-categories in the metabolic function were "binding" and "catalytic activity". The overall cellular component category was less represented in the up-regulated DEGs (Figure 5.4) and more pronounced in the down-regulated DEGs with the sub-categories: "membrane", "membrane part", "cell", "cell part" and "organelle" (Figure 5.5).

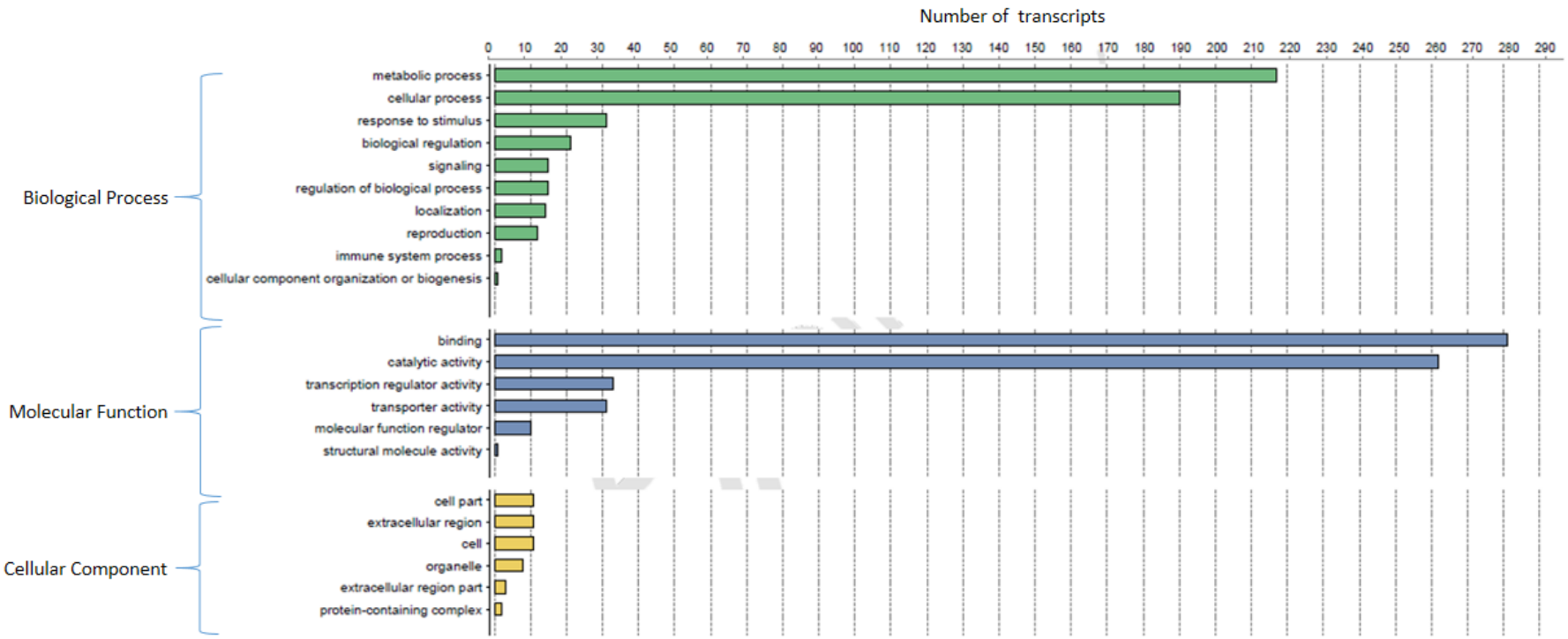


**Figure 5.4 Gene ontology (GO) classifications (level 2) of up-regulated transcripts between normal and crinkle dwarf phenotypes.** The biological process in green, molecular function in blue and cellular component in yellow.

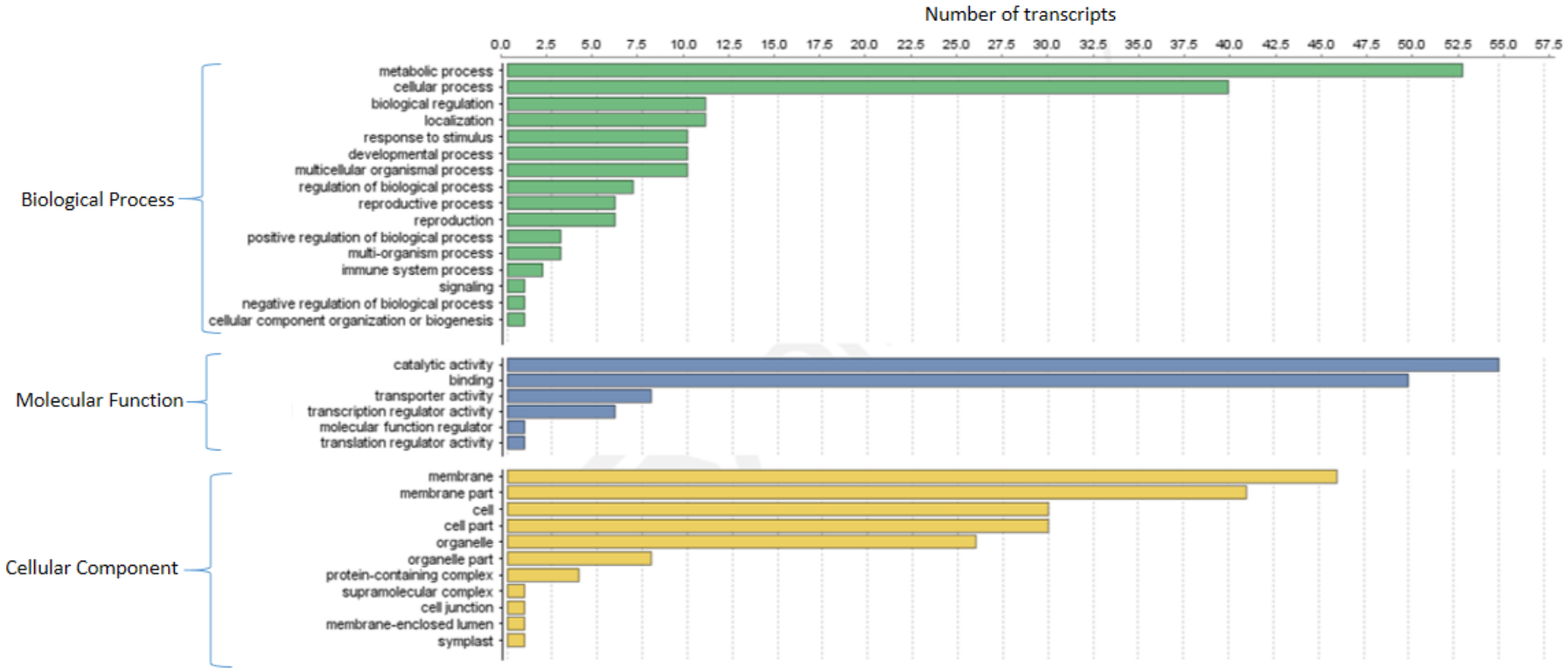


**Figure 5.5 Gene ontology (GO) classifications (level 2) of down-regulated transcripts between normal and crinkle dwarf phenotypes.** The biological process in green, molecular function in blue and cellular component in yellow.

***Clusters of Orthologous Groups (COG)***

The differentially expressed transcripts (763 up- and 158 down-regulated) were aligned against the COG database for functional prediction and classification to improve the understanding of biological functions pertaining to crinkle dwarf phenotype. The classification includes coding of proteins against orthologous genes based on the evolutionary relationships of bacteria, algae, and eukaryotes (Huerta-Cepas *et al.*, 2019).

Of the total, 340 (44.56 %) up-regulated and 42 (26.58 %) down-regulated transcripts were classified into 25 COG functional categories, and were subdivided based on A-Z coding (Figure 5.6; Supplementary Table 5.3). The top five categories in the up-regulated DEGs were, "Function unknown (S)" which represented the largest group (207, 27.1 %), followed by "Signal transduction mechanisms (T) (144, 18.9 %), "Posttranslational modification, protein turnover, chaperones (O)" (58, 7.6 %), "Transcription (K)" (58, 7.6 %) and "Secondary metabolites biosynthesis, transport and catabolism (Q)" (57, 7.5 %). In the down-regulated DEGs, ''Function unknown (S)" represented the largest group (32; 20.2 %), followed by the "Transcription (K)" (9, 5.7 %), "Lipid transport and metabolism (I)" (7, 4.4%), "Replication, recombination and repair (L)" (4, 2.5 %) and, "Signal transduction mechanisms (T)" (4, 2.5 %) Figure 5.6; Supplementary Table 5.3).

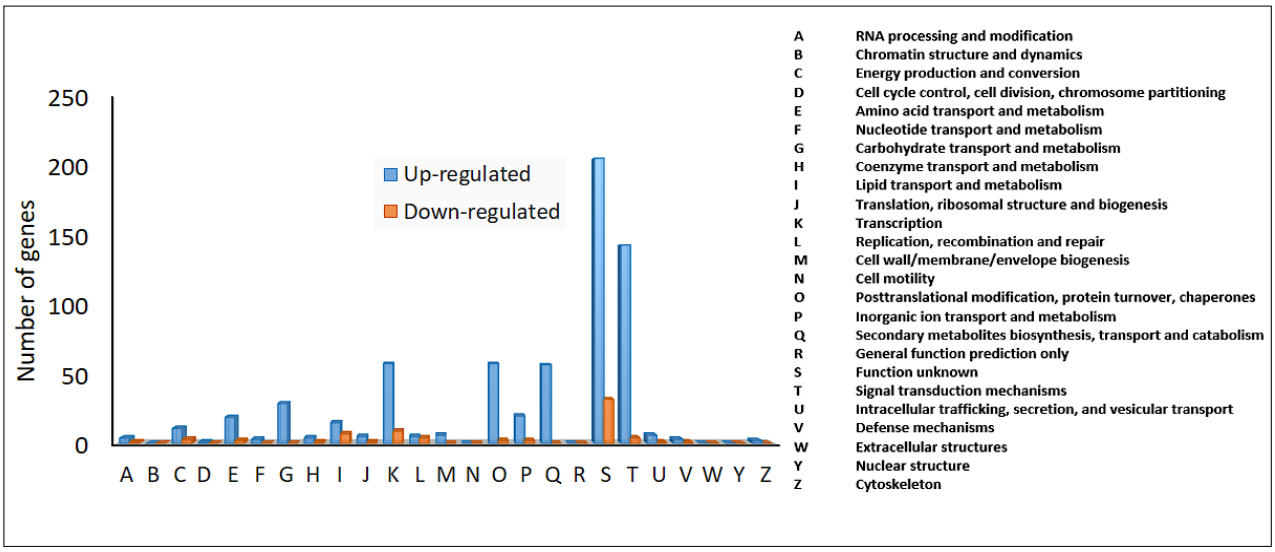


**Figure 5.6 COG distribution of up- and down-regulated DEGs between normal and crinkle dwarf phenotype.**

### Kyoto Encyclopedia of Genes and Genomes (KEGG)

To broaden the understanding and the emphasis of the biological functions, the 340 and 42 differential expressed transcripts were annotated against the KEGG pathway database. The KEGG pathway database is a widely accepted source for biological pathway maps. In total, 41 and 20 up-regulated and down-regulated transcripts were assigned to 23 (Supplementary Table 5.4) and 20 (Supplementary Table 5.5) KEGG pathways, respectively.

The differentially expressed transcripts annotated with KEGG terms were categorised into two major categories: metabolism, and biosynthesis pathways. Of these, five metabolism and two biosynthesis pathways were common in both up- and down-regulated transcripts, though regulated by different enzymes with the exception of nitrogen metabolism (Supplementary Table 5.4; Supplementary Table 5.5). In the metabolism category, the pathways were classified into “cysteine and methionine metabolism”, “purine metabolism”, “thiamine metabolism”, “riboflavin metabolism” and “nitrogen metabolism”. The biosynthesis pathways were “phenylpropanoid biosynthesis” and biosynthesis of antibiotics. The secary metabolites also consisted of variable

subcategories amongst the up- and down-regulated transcripts (Supplementary Table 5.4; Supplementary Table 5.5).

## 5.5 Discussion

### 5.5.1 Sequencing and Differential Expressed Genes (DEGs)

The tissues selected for the current study are believed to be key determinants in plant development (Barthélémy and Caraglio, 2007; Poethig, 2010; Xu *et al.*, 2016). The use of meristematic tissues including shoot apical meristems (SAMs) and young leaves has proven successful in transcriptome-based studies of apple (Krost *et al.*, 2012) and pear (Ou *et al.*, 2015). The transcriptome profiling data acquired from the pools (apical buds and young leaves) of normal and crinkle phenotypes served as a valuable resource in exploring the underlying genetic mechanisms associated with crinkle dwarf phenotype, particularly in the cultivars of 'McIntosh' and 'M.1'.

The high-quality cleaned reads were successfully mapped to the apple reference transcriptome, with a high average mapping percentage of 87.80% (Table 5.1; Supplementary Table 5.1). In total, 921 transcripts were determined to be expressed differentially at a significant selection criterion based on log2foldchange > +1.5 and < -1.5, p-adjusted < 0.05 and FDR < 0.01.

### 5.5.2 Genes and proteins associated with peroxidase EC:1.11.1.7

The phenylpropanoid pathway is one of the most important metabolic pathways in plants (Naoumkina *et al.*, 2010). In this study, lactoperoxidase (EC:1.11.1.7) synthesized from the pathway of phenylpropanoid biosynthesis was the most abundantly expressed amongst all the DEGs. Plant peroxidases (EC:1.11.1.7), often designated as class III peroxidases (abbreviated as *Prx*), are plant-specific, heme-containing glycoproteins and belong to a large multigenic family (Almagro *et al.*, 2009; Francoz *et al.*, 2015; Shigeto and Tsutsumi, 2016). In the regular peroxidative cycle, peroxidases catalyse the reduction of $H_2O_2$ to various donor molecules (Hiraga *et al.*, 2001).

Peroxidases (Prx) are involved in a wide range of physiological processes, such as lignification, suberisation, auxin catabolism, wound healing and defense mechanisms against pathogen infection

(Kawano, 2003). The plant Prxs in the apoplastic space are often involved in reactive oxygen species (ROS) generation. These effects lead to growth retardation, metabolic disturbances and oxidative stress. Plants may tolerate and adapt to these stressors with different mechanisms including changed leaf architecture, osmotic adjustment, ion exclusion and compartmentalization and a more efficient ROS scavenging systems. Depending on the strength and duration of these stresses as well as the plant's genetic composition, it can be a matter of life or death since there is always the so-called "stability limit", which when exceeded, leads to death in the organism.

Additionally, peroxidase are also implicated in wound-healing through its catalytic role in the cross-linking of pectin and structural proteins in the cell wall and in the polymerisation of the phenolic monomers of lignin (Lagrimini *et al.*, 1997). In tomato (*Lycopersicon esculentum*), peroxidase gene (*TPX1)* was induced in vascular tissue of aerial parts after wounding (Mohan *et al.*, 1993*b*; Botella *et al.*, 1994). In an additional study of tomato, two peroxidase genes (*tap1*, *tap2*) were expressed as a result of wounding in fruit, leaf and stem tissues. These homologous genes code for anionic peroxidases that are suggested to cause polymerisation of the phenolic residues into cell wall polymers in wounded and pathogen-infected tissues (Mohan *et al.*, 1993*a*,*b*).

Moreover, the synthesis of lignin and suberin as well as the cross-linking of these substances with other wall polymers is also catalyzed by cell-wall peroxidases (Espelie *et al.*, 1986). Therefore, peroxidases are key enzymes for cell wall reinforcement and may also play an important role as intermediaries in signal cascades triggered by external stimuli to the cell wall (Passardi *et al.*, 2005; Francoz *et al.*, 2015).

### 5.5.3 Genes and proteins associated with glutathione S-transferase (GSTs, EC 2.5.1.18)

Glutathione transferases, known as glutathione *S*-transferases (GSTs; EC 2.5.1.18) are a superfamily of multifunctional enzymes which functions in detoxification of a wide range of endobiotic and xenobiotic compounds. In this capacity, GST catalyzes the conjugation of the tri-peptide glutathione (GSH; γ-Glu–Cys–Gly), rendering the substrate less toxic (Marrs, 1996; Dixon and Edwards, 2010; Cummins *et al.*, 2011). Moreover, GST may exhibit a glutathione-peroxidase activity associated with the

reduction of hydrogen peroxide (e.g. oxidative stress tolerance) (Wagner *et al.*, 2002) or function as ligand-binding protein (Wagner *et al.*, 2002; Dixon and Edwards, 2010)

In plants, GST has been implicated to respond to a wide range of factors including phytohormones, herbicides, heavy metals, wounding, pathogen attack, ozone, and hydrogen peroxide (Marrs, 1996; Eaton and Bammler, 1999; Gullner *et al.*, 2018). In *Arabidopsis*, GST have been shown to be induced by phytohormones: SA, auxin, ethylene, cytokinin, methyl jasmonate, and brassinosteroid (Wagner *et al.*, 2002; Sappl *et al.*, 2004; Deng *et al.*, 2007; Dixon *et al.*, 2009). Marrs, (1996) catergorise GST as a secary defense enzyme. In yeast, it was demonstrated that a tomato (*Lycopersicon esculentum*) *GST* gene (*Bak*-stimulated apostosis) inhibited cell death and elevated resistance to $H_2O_2$ stimulated stress (i.e. oxidative stress-tolerance) (Kampranis *et al.*, 2000). In pear, a GST gene (*PpGST1*) was induced in diseased fruit cells, suggesting their role in disease resistance during fruiting and senescence (Shi *et al.*, 2014). In *Arabidopsis*, a GST gene (*GSTF8)* is used as a marker for early stress and defense responses (Perl-Treves *et al.*, 2004; Thatcher *et al.*, 2007, 2015; Gleason *et al.*, 2011). These studies demonstrated the multifunctionality of GST in regulating plant growth and development, and its involvement in defense-related responses.

### 5.5.4 Genes and proteins associated with amino acids

Proline is one of the amino acids whose metabolism impacts multicellular functions (Cecchini *et al.*, 2011). For many years, the capacity to accumulate proline has been correlated with stress tolerance (Liang *et al.*, 2013). Proline acts as an osmolyte, a ROS scavenger, and a molecular chaperone stabilizing the structure of proteins, thereby protecting cells from damage caused by stress (Zhang and Becker, 2015). In *Caenorhabditis elegans*, proline has been implicated in modulating innate immunity by governing ROS homeostasiss and subsequent activation of a transcription factor (SKN-1) regulating xenobiotic stress response and pathogen defense (Tang and Pang, 2016).

In *Colletotrichum trifolii*, a fungal pathogen of alfalfa, proline acts as a potent ROS scavenger associated with prevention of apoptotic-like program cell death (PCD) (Chen and Dickman, 2005). Moreover, proline protective mechanisms during osmotic stress have been proposed to involve the stabilization of proteins and antioxidant enzymes, direct scavenging of ROS, balance of intracellular

redox homeostasis, and cellular signaling promotion (Cecchini *et al.*, 2011; Liang *et al.*, 2013). In *Arabidopsis*, proline has been implicated to potentiates the oxidative burst and cell death associated to the hypersensitive responses (HR). Interestingly, activation of ProDH can also produce harmful effects in other organisms, suggesting that the enzyme may play a conserved role in the control of cell death. Therefore, the upregulation of proline dehydrogenase (EC:1.2.1.41) in the present study suggest the relation of crinkle dwarf phenotype to auto-immune response.

### 5.5.5 Pathogenesis-related (PR) genes: chitinase and pectinase

#### *ChiC*

Chitin is the polymer of N-acetylglucosamine and constitutes the cell walls of a wide range of fungi. Chitin oligosaccharides are representative fungal microbe-associated molecular patterns (MAMPs) that elicit various defense responses in many plant species (Kombrink et al. 2011).

Chitin is the sec most abundant polysaccharide after cellulose. It is a vital structural component of the fungal cell wall but not for plants. In plants, fungi are recognized through MAMPs. Therefore, chitin is a fungal microbe-associated molecular pattern (Cao *et al*., 2014). Chitinases are hydrolytic enzymes catalyzing the hydrolysis of β-1,4-linkage of the *N*-acetylglucosamine polymer of chitin (Pusztahelyi, 2018). Plant chitinases have molecular weight varying from 25 to 40 kDa with both acidic and basic isoforms constitutively present in stems, seeds, flowers and tubers. Chitinase (EC:3.2.1.14) belongs to the families of PR proteins: PR-3, PR-4, PR-8, and PR-11, that respond to pathogen attacks. Chitinases play a role in plant defense by directly or indirectly inhibiting the hyphae growth which invades the intercellular space or release fungal elicitors. In *Arabidopsis*, the transgenic plants overexpressing chitinases confers increase resistance to pathogen attack (Takenaka *et al.*, 2009). In this study, ChiC (EC:3.2.1.14) was up-regulated, suggesting crinkle dwarf phenotype elicits a response similar to fungal pathogenic attack.

#### *Pectin*

Pectinesterase also known as pectin-demethylase (EC:3.1.1.11) was up-regulated. Pectin are categorised under pathogenesis-related proteins. They are involved in enhancing cell-to-cell adhesion,

cell elongation, porosity of the wall, disease resistance and ultimately plant growth and development (Liu *et al.*, 2018).

### 5.5.6 Genes and proteins associated with cutin, suberine and wax biosynthesis EC:2.3.2.188

A down-regulation was noted for cutin, suberine and wax biosynthesis encoding a suberin-associated omega-hydroxypalmitate O-feruloyl transferase (EC:2.3.1.188), O-feruloyl transferase.

Suberin is a cell wall-linked polymeric barrier that plays a critical role in the survival of plants by protecting them against various biotic and abiotic stresses. It primarily acts as a protective hydrophobic barrier to control the movement of water, gases, and solutes, and also contributes to the strengthening of the cell wall (Franke *et al.*, 2005; Franke and Schreiber, 2007; Ranathunge *et al.*, 2011). Suberin comprised of polyaliphatic and polyaromatic domains. The predominant aliphatic components of suberin are ω-hydroxy fatty acids, α,ω-dicarboxylic acids, glycerol and ferulate. The major phenolic components are *p*-hydroxycinnamic acids, especially ferulic acid (Serra *et al.*, 2010; Ranathunge *et al.*, 2011; Delude *et al.*, 2016).

In *Arabidopsis*, aliphatic suberin feruloyl transferase (At5g41040), a HXXXD-type acyltransferase (BAHD family), is responsible for the incorporation of ferulate into aliphatic suberin (Molina *et al.*, 2009; Serra *et al.*, 2010; Vishwanath *et al.*, 2013). Feruloyl transferase activity utilising feruloyl-CoA as the acyl donor and *ω*-OHFA as acyl acceptor has been demonstrated in extracts from potato tubers and tobacco cell suspensions (Lotfy and Javelle, 1994; Lotfy *et al.*, 1996). The activity of a ferulate has been reported in the wound-healing potato-tuber discs and in cultures of tobacco (*Nicotiana* spp.) cells (Lotfy and Javelle, 1994; Lotfy *et al.*, 1996). In potato (*Solanum tuberosum*), a gene encoding a fatty ω -hydroxyacid/fatty alcohol hydroxycinnamoyl (feruloyl transferase) was down-regulated, and had significant effects on the anatomy, sealing properties and maturation of the periderm. Moreover, the tuber skin became thicker and russeted (rough scabbed skin), increased water loss, and impaired skin maturation (Serra *et al.*, 2010). Serra *et al*. (2010), also mentioned that a russeted skin is commonly associated to the scab disease caused by a bacteria (*Streptomyces* spp.), which secretes a toxin (thaxtomin). In this study, down-regulation of feruloyl transferase may be associated with the crinkled

leaves (crinkle dwarf phenotype) which are brittle in texture and having a russetted midrib on the underside of the leaves.

### 5.5.7 Genes and proteins associated with serine/threonine phosphatase EC:3.1.3.16

The protein serine/ threonine phosphatase is a form of phosphoprotein phosphatase that acts upon phosphorylated serine/threonine residues. Serine and threonine phosphates are amino acids which have similar side-chain compositions that contain a hydroxyl group which enables phosphorylation by serine/threonine protein kinases (Brautigan, 2013; Ghosh *et al.*, 2014). The reversible phosphorylation of proteins is accomplished by the opposing activities of kinase and phosphatases. The addition and removal of phosphate groups regulates many cellular pathways including cell proliferation, programmed cell death (apoptosis), embryonic development, and cell differentiation (Afzal *et al.*, 2008; Uhrig *et al.*, 2013; Ghosh *et al.*, 2014).

In the present study, phosphatase (EC:3.1.3.16), a serine/threonine phosphatase 2B catalytic subunit (PP2B; also known as calcineurin) was up-regulated. The calcineurin is a calcium dependent, calmodulin (CaM)-stimulated protein phosphatase (an essential calcium transducer) (Cheval *et al.*, 2013; Nygren and Scott, 2016). The specific name of the protein, PP2B, is based on the pathways "Th1 and Th2 cell differentiation", PD-L1 expression and PD-1 checkpoint pathway in cancer" and "T-cell receptor signaling pathway" (Supplementary Table 5.6.). The PP2B plays an essential role in the transduction of intracellular calcium ($Ca^{2+}$) mediated signals. On the other hand, calcium participates in the structural integrity of the cell wall and membrane system. Calcium also acts as an intracellular regulator in many aspects of plant growth and development including stress response (Galon *et al.*, 2010). The PP2B consist of two subunits: a catalytic subunit (calcineurin A; CnaA) and a regulatory subunit (calcineurin B; CnaB) (Batistič and Kudla, 2009; Juvvadi *et al.*, 2014; Nygren and Scott, 2016). Calcineurin is an essential gene involved in hyphal growth (i.e. regulates fungal stress responses, morphogenesis and pathogenesis) and maintanance of $Ca^{2+}$ gradients. Fungal pathogens utilize the calcineurin pathway to survive in the host environment and cause life-threatening infections (Juvvadi *et al*. 2014). Therefore, the current results also suggests that the metabolism leading to crinkle dwarf

phenotype may be similar as to that of fungal pathogenic attack, however, further validations are necessary.

An *Arabidopsis* calcium/calmodulin-binding receptor-like cytoplasmic kinase (AtCRCK1) shows high homology in the kinase domain to that of serine/threonine receptor-like kinase. The expression of AtCRCK1 is induced in response to cold and salt stress including abscisic acid and $H_2O_2$, which is an indication of the presence of calcium/calmodium-regulated receptor-like cytoplasmic kinase (Poovaiah *et al.*, 2013). These results in turn suggest that calcium/calmodium-regulated protein phosphorylation plays a major role in stress signal transduction.

Interestingly, in maize *(Zea mays* L.), *CRINKLY4 (CR4)* gene encodes a serine/threonine receptor-like kinase (RLK) that controls an array of developmental processes in the plant and endosperm (Becraft *et al.*, 1996, 2001; Bowles *et al.*, 2000). CR4 is expressed in the growing regions of the shoot, particularly in the shoot apical meristem (SAM) and lateral organ primordia (Bowles *et al.*, 2000; Becraft *et al.*, 2001). Loss of function *cr4* mutant affects leaf epidermis differentiation (Becraft *et al.*, 1996). The maize *cr4* mutant plants are short in stature with crinkled leaves showing a graft-like tissue fusions, as a result of wrong cell fate specification of the endosperm (Becraft and Asuncion-Crabb, 2000; Bowles *et al.*, 2000; Jin *et al.*, 2000). Importantly, CR4 encodes an RLK with two different domains, one, is similar to the ligand-binding domain of mammalian tumour necrosis factor receptor (TNFR) (Becraft *et al.*, 1996). The sec, extracellular domain consists of seven repeat regions of about 39 amino acids, called crinkly repeats. The crinkly repeats are hypothesized to form a regulator of chromosome condensation 1 (RCC1)-like propeller structure. Both the domains are thought to participate in protein-protein interactions (Becraft *et al.*, 1996; Cao *et al.*, 2005; Gifford *et al.*, 2005).

Similarly, *Arabidopsis* encodes an ortholog of the maize CR4, named ACR4, and four crinkly4-related (CRR) proteins: AtCRR1, AtCRR2, AtCRR3 and AtCRK1. contains a family of five receptor-like kinase RLKs related to CR4, and named ACR4 (Tanaka *et al.*, 2002; Gifford *et al.*, 2003; Cao *et al.*, 2005; Czyzewicz *et al.*, 2016). The ACR4 possesses an extracellular ligand-binding domain, a transmembrane helix, and an intracellular domain that contains the juxtamembrane and the C-terminal subdomains, which flank the core kinase domain with serine/threonine kinase activity(Gifford *et al.*,

2005; Czyzewicz *et al.*, 2016). The *Arabidopsis acr4* mutant has been shown to affect the differentiation of leaf epidermal cells, suggesting a similar role for ACR4 and CR4 in the differentiation of leaf (Tanaka *et al.*, 2002; Gifford *et al.*, 2003; Cao *et al.*, 2005).

Another gene in *Arabidopsis*, abnormal leaf shape1 (ALE1), encodes a putative subtilism-like serine protease and is expressed in the endosperm cells surrounding the developmental embyros (Tanaka *et al.*, 2001; Watanabe *et al.*, 2004). The mutant (*ale1*) causes impaired formation of the cuticle (Tanaka *et al.*, 2001; Watanabe *et al.*, 2004). On the contrast, *Arabidopsis*,an abnormal leaf shape2 (ALE2) acts with the ACR4 during embryo development (Tanaka *et al.*, 2007). The mutant (*ale2*) has similar epidermal defects to that of maize (*cr4*) and *Arabidopsis* (*acr4*) mutants which are characterised by having crinkly leaves, defective cuticle and abnormal ovule development. Interestingly, the double mutant *acr4 ale2* resemble *ale2* mutants, indicative of the involvement of the same RLK signaling pathway (Tanaka *et al.*, 2002, 2007; Watanabe *et al.*, 2004).

Moreover, in *Arabidopsis*, a protein serine/threonine phosphate 2A-3 (PP2A-3), which is a catalytic subunit of PP2A holoenzymes that functions in regulating formative cell divisions has been described as the substrate for ACR4 at the phosphorylation level (i.e. PP2A dephosphorylates ACR4). Taken together, these findings suggest that crinkle dwarf phenotype may also be regulated by the serine/threonine phosphatase which encodes for a RLK possible PP2A or by the both the PP2A and PP2B. However, further studies are required to unveil the specificity of the PP2 signaling pathway in relation to crinkle dwarf phenotype in apple.

### 5.5.8 Genes and proteins associated with phytohormones

Plant hormones are not only essential for plant growth and development but also plays a central roles in triggering the plant immune signaling network Plants typically respond to pathogen attack with a complex scenario of sequential, antagonistic, or synergistic action (i.e. hormonal cross-talk) of different hormone signals leading to defense gene expression (Robert-Seilaniantz *et al.*, 2010; Pieterse *et al.*, 2012).. This complexity requires a tight regulation as individual hormone signaling pathways are intertwined and any regulatory disorder would impair this network and thus disturb growth or

immunity. The phytohormones salicylic acid (SA), jasmonic acid (JA) and ethylene (ET)) play a key role in the regulation of disease signaling pathways (Thaler *et al.*, 2012). Salicylic acid is perceived effective against biotrophic pathogens (i.e. feeding on living plant tissues), whereas cell death provoking necrotrophic pathogens are usually deterred by JA and ET pathways (Bari and Jones, 2009). De Vleeschuwer *et al*. (2014) stated that the SA and JA/ET hormonal pathways interact in an antagonistic manner i.e. having opposite roles, suggesting their major role in plant immunity. In this study, alpha-linolenic, a precursor of jasmonic acid was up-regulated. Moreover, other phytohormones, including canonical growth hormones auxin, brassinosteroids (BR), gibberellins (GA), cytokinins (CY) were also found to affect immune signaling.

### 5.5.9 Genes and proteins associated with GA2ox (EC:1.14.11.13)

Gibberellins are synthesized by a terpenoid pathway in plastids and modified in the endoplasmic reticulum and cytosol (Otani *et al.*, 2013). The last step of the GA biosynthesis is catalysed by three groups of oxygenases and are localized in the cytosol; GA 20-oxidase (*GA20ox*), GA 3-oxidase (*GA3ox*) and GA 2-oxidase (*GA2ox*) (Hedden and Phillips, 2000; Olszewski *et al.*, 2002; Sun and Gubler, 2004). However, *GA20ox* and *GA3ox* produces bioactive GAs while *GA2ox* deactivates bioactive GAs. In particular, *GA2ox*s catalyses the deactivation of bioactive GAs or its precursors through 2β-hydroxylation reaction, resulting in biologically inactive products that cannot be restored exogenously (Olszewski *et al.*, 2002; Daviere and Achard, 2013). Gibberellin 2β-hydroxylase activity is reported to be abundant in seeds during the later stages of maturation (Albone *et al.*, 1984; Thomas *et al.*, 1999; Singh *et al.*, 2002; Solfanelli *et al.*, 2005).

The diterpenoid biosynthesis 2-beta-dioxygenase pathway has been suggested to be a major mechanism for GA inactivation, and a schematic GA-pathway.In previous studies, overexpression of *GA2ox* in *Arabidopsis* (Rieu *et al.*, 2008), tobacco (*Nicotiana tabacum*) (Schomburg *et al.*, 2003), *Tricyrtis* species (Otani *et al.*, 2013) and in rice (*Oryza sativa*) (Huang *et al.*, 2010) resulted in dwarf phenotypes. In contrast, the suppression of *GA2ox* expression was demonstrated in the pea (*Pisum sativum*) *slender* mutant which resulted in hyper-elongation, characterised by a tall and slender phenotype (Martin *et al.*, 1999). In the present study, the 2-beta-dioxygenase is down-regulated,

suggesting a suppression of *GA2ox*, and the possibility of increased levels of gibberellin in the crinkle dwarf phenotype. Dwarf growth habits which occurs as a result of mutations in the GA biosynthesis or signaling can exhibit dwarf phenotypes (Peng *et al.*, 1997; Griffiths *et al.*, 2006; Leubner-metzger, 2011). Our results suggest that crinkle dwarf phenotype may be gibberellin-insensitive i.e. crinkle dwarf may not be regulated by GA, but may occur as a result of mutation(s) in the GA biosynthesis or signaling pathway.

Additionally, GA signaling is mediated via gibberellin-insensitive dwarf1 (GID1) and DELLA proteins (Sun, 2011). DELLA transcription factors are repressors of GA signaling (Sun, 2008, 2011). DELLA proteins play important roles not only in GA signaling, but also in the signaling of other plant hormones and environmental cues that are important for plant growth and development (Komorisono, 2005; Cheng *et al.*, 2019).

### 5.5.10 Genes and proteins associated with monooxygenases (EC:1.14.13.8) and tryptophan metabolism, O-methyltransferase (EC:2.1.1.4)

Flavin-containing monooxygenases (FMOs, EC 1.14.13.8) are a group of enzymes that catalyzes the oxygenation of nitrogen and sulfur in small organic molecules (Dai *et al.*, 2013). Plants use FMOs to synthesize signaling molecules that play essential roles in many aspects of plant growth and development (Cheng *et al.*, 2006, 2007).

In *Arabidopsis*, an enzyme, YUCCA1 (YUC1) flavin-containing monooxygenases is involved in auxin biosynthesis and controls the formation of floral organs and vascular tissues (Dai *et al.*, 2013). YUCCA1 is a key enzyme that catalyzes a rate-limit step in a tryptophan-dependent auxin biosynthesis pathway (Cheng *et al.*, 2006). Overexpression of YUC genes in *Arabidopsis* leads to auxin overproduction. In contrast, inactivation of YUC genes causes defects in embryogenesis, seedling growth, vascular and floral development (Cheng *et al*., 2006, 2007; Wang *et al*., 2011). Auxin are phytohormones that plays a vital role in leaf development, particularly in phyllotaxis of leaf formation (Reinhardt *et al*., 2000). During leaf development, the formation of leaf adaxial-abaxial polarity at the primordium stage is crucial for subsequent leaf expansion (Qi *et al*., 2014; Dong and Huang, 2018). In

this study, flavin monooxygenases (FMOs, EC 1.14.13.8) is up-regulated, and may have an effect on the crinkled leaf development, however, further studies on auxin-YUCCA are needed.

Interestingly, in the current study, "tryptophan metabolism" O-methyltransferase (EC:2.1.1.4) is down-regulated. O-methyltransferase is an enzyme that catalyzes the final reaction in melatonin biosynthesis. Melatonin (N-acety-5-methoxytryptamine), is an indoleamine molecule that is widely found in animals and plants (Hattori *et al*., 1995). Melatonin is regarded as a candidate phytohormone "phytomelatonin" and functions as a metabolite with pleiotropic biological activities including biological rhythms, stress tolerance, plant growth and development (Arnao and Hernández-Ruiz, 2015; Wang *et al*., 2018). Melatonin has been found in different plant species and various organs, such as roots, stems, leaves, fruits, and seeds (Hernández-ruiz, 2006).

In plants, melatonin has a similar biological activity as auxin and shares the same initial biosynthetic precursor, tryptophan. As such, melatonin functions as an indole-3-acetic acid (IAA) like hormone. The biosynthetic pathway of melatonin is synthesized from tryptophan by four key enzymes: tryptophan decarboxylase (TDC), tryptamine 5-hydroxylase (T5H), serotonin *N*-acetyltransferase (SNAT), and *N*-acetylserotonin *O*-methyltransferase (HIOMT/ASMT).

The effect of melatonin on plant growth development has been studied extensively in *Arabidopsis*. A high dose of melatonin inhibited root growth in *Arabidopsis* seedlings by reducing root meristem size (Wang *et al.*, 2016). In another study, high dosage of melatonin-mediated protein suppressed floral transition in *Arabidopsis* (Shi *et al.*, 2016). Higher concentration of melatonin has also been reported to have a negative effect in leaf development. In *Arabidopsis*, overexpression of melatonin was reported to regulate leaf development by suppressing cell proliferation and cell expansion causing defects during endoreduplication. Consequently, affecting regulated transcriptional levels of cell cycle and ribosomal key genes (Wang *et al.*, 2017).

In the present study, it is likely that the up-regulation of the flavin-containing monooxygenases (EC 1.14.13.8) and down-regulation of tryptophan metabolism, O-methyltransferase (EC:2.1.1.4), respectively, may have a biological impact on the development and therefore manifestation of the

crinkled leaves. However, this assumption would require a more directed experimental approach to investigate the effect of auxin-YUC-melatonin on crinkled leaf development.

## 5.6 Conclusion and Future Work

Transcriptome profiling of differentially expressed genes between normal and crinkle dwarf phenotypes unveiled defense signaling and stress-related proteins/enzymes that may have an effect in regulating crinkle dwarf phenotype. A cross-talk including ROS-associated (peroxidase, glutathione S-transferase), serine/threonine phosphatase receptor-like kinase (PP2), pathogenic-related proteins (chitinase and pectin) together with cutin, suberin and wax related genes, may be possible causal enzymes/proteins related to the molecular mechanism underlying the crinkle dwarf phenotype. Although the experimental approach followed here provided insight to the possible molecular mechanisms associated with crinkle dwarf phenotype, particularly in cultivars of 'McIntosh' and 'M.1', further work is needed to validate these findings. Currently, additional work is being conducted on the same RNA samples, however, the developmental stages (Figure 5.1) are being sequenced independently, with the aim to investigate transcriptomes between and within developmental stages. Furthermore, quantitative reverse transcription PCR (RT-qPCR) will be conducted on several of the most important genes identified by the transcriptome analysis in order to show independently that the same trends in gene expression (i.e. up- or down-regulation) are observed. In future, these findings could be of practical implications and contribute to knowledge needed in designing novel breeding strategies whereby raising dwarf seedling associated with crinkle leaves can be avoided.

## 5.7 Supplementary Data

Please see back of document

Supplementary Table 5.1 Raw, clean and mapped reads between each Pools of normal and crinkle dwarf phenotypes.

Supplementary Table 5.2 Concentrations of constructed cDNA- libraries.

Supplementary Table 5.3 Up-regulated differential expressed genes between normal and crinkle dwarf phenotypes at log2foldchange >+1.5 and <-1.5, p-adjusted <0.05, FDR <0.01.

Supplementary Table 5.4 Down-regulated differential expressed genes between normal and crinkle dwarf phenotypes at log2foldchange >+1.5 and <-1.5, p-adjusted <0.05, FDR <0.01.

Supplementary Table 5.5 COG functional classification of up- and down-regulated genes between normal and crinkle dwarf phenotypes.

Supplementary Table 5.6 Enzyme mapping and annotations according to KEGG for up-regulated DEGs (excel file).

Supplementary Table 5.7 Enzyme mapping and annotations according to KEGG for down-regulated DEGs (excel file).

**Supplementary Figures**

Supplementary Figure 5.1 KEGG pathway of diterpenoid biosynthesis illustrating the 2-beta-dioxygenase pathway (*GA2ox*) (EC:1.14.11.13). The GA catabolic *GA2ox* (EC:1.14.11.13) is highlighted in yellow boxes. The EC number represents the enzyme commission code.

## 5.8 References

**Afzal AJ, Wood AJ, Lightfoot DA**. 2008. Plant receptor-like serine threonine kinases: Roles in signaling and plant defense. Molecular Plant-Microbe Interactions **21**, 507–517.

**Albone KS, Gaskin P, Macmillan J, Sponsel V**. 1984. Identification and localization of gibberellins in maturing seeds of the cucurbit *Sechium edule*, and a comparison between this cucurbit and the legume *Phaseolus coccineus*. Planta **162**, 560–565.

**Alcázar R, García A V., Kronholm I, De Meaux J, Koornneef M, Parker JE, Reymond M**. 2010. Natural variation at strubbelig receptor kinase 3 drives immune-triggered incompatibilities between *Arabidopsis thaliana* accessions. Nature Genetics **42**, 1135–1139.

**Alcázar R, García A V., Parker JE, Reymond M**. 2009. Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. Proceedings of the National Academy of Sciences of the United States of America **106**, 334–339.

**Almagro L, Gómez Ros L V., Belchi-Navarro S, Bru R, Ros Barceló A, Pedreño MA**. 2009. Class III peroxidases in plant defence reactions. Journal of Experimental Botany **60**, 377–390.

**Anders S, Huber W**. 2010. Differential expression analysis for sequence count data. Genome Biology **11**, R106.

**Anders S, Pyl PT, Huber W**. 2015. HTSeq - a Python framework to work with high-throughput sequencing data. Bioinformatics **31**, 166–169.

**Arnao MB, Hernández-Ruiz J**. 2015. Functions of melatonin in plants: a review. Journal of Pineal Research **59**, 133–150.

**Atkinson CJ, Brennan RM, Jones HG**. 2013. Declining chilling and its impact on temperate perennial crops. Environmental and Experimental Botany **91**, 48–62.

**Bai Y, Dougherty L, Xu K**. 2014. Towards an improved apple reference transcriptome using RNA-seq. Molecular Genetics and Genomics **289**, 427–438.

**Bai S, Saito T, Sakamoto D, Ito A, Fujii H, Moriguchi T**. 2013. Transcriptome analysis of japanese pear (*Pyrus pyrifolia* Nakai) flower buds transitioning through endodormancy. Plant and Cell Physiology **54**, 1132–1151.

**Bari R, Jones JDG**. 2009. Role of plant hormones in plant defence responses. Plant Molecular Biology **69**, 473–488.

**Barthélémy D, Caraglio Y**. 2007. Plant architecture: A dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. Annals of Botany **99**, 375–407.

**Batistič O, Kudla J**. 2009. Plant calcineurin B-like proteins and their interacting protein kinases. Biochimica et Biophysica Acta **1793**, 985–992.

**Becraft PW, Asuncion-Crabb Y**. 2000. Positional cues specify and maitain aleurone cell fate in maize endosperm development. Development **127**, 4039–4048.

**Becraft PW, Kang SH, Suh SG**. 2001. The maize CRINKLY4 receptor kinase controls a cell-autonomous differentiation response. Plant Physiology **127**, 486–496.

**Becraft PW, Stinard PS, McCarty DR**. 1996. Crinkly4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science **273**, 1406–1409.

**Benjamini Y, Hochberg Y**. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society **57**, 289–300.

**Bindschedler L V, Dewdney J, Blee K a, *et al.*** 2006. Peroxidase-dependent apoplastic oxidative burst in <i)*Arabidopsis*</i> required for pathogen resistance. Plant Journal **47**, 851–863.

**Bolger AM, Lohse M, Usadel B**. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30**, 2114–2120.

**Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D**. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biology **5**, 1962–1972.

**Bomblies K, Weigel D**. 2007. Hybrid necrosis: Autoimmunity as a potential gene-flow barrier in plant species. Nature Reviews Genetics **8**, 382–393.

**Botella MA, Quesada MA, Medina MI, Pliego F, Valpuesta V**. 1994. Induction of a tomato peroxidase gene in vascular tissue. FEBS Letters **347**, 195–198.

**Bowles J, Bullejos M, Koopman P**. 2000. The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. Genesis **27**, 104–116.

**Brautigan DL**. 2013. Protein Ser/Thr phosphatases - The ugly ducklings of cell signaling. FEBS Journal **280**, 324–345.

**Burke JM, Arnold ML**. 2001. Genetics and the fitness of hybrids. Annual Review of Genetics **35**, 31–52.

**Cao X, Li K, Suh SG, Guo T, Becraft PW**. 2005. Molecular analysis of the CRINKLY4 gene family in *Arabidopsis* thaliana. Planta **220**, 645–657.

**Cecchini N, Monteoliva M, Alarez M**. 2011. Proline dehydrogenase is a positive regulator of cell death in different kingdoms. Plant Signal and Behavior **6**, 1195–1197.

**Chen C, Dickman MB**. 2005. Proline suppresses apoptosis in the fungal pathogen *Colletotrichum trifolii*. Proceedings of the National Academy of Sciences of the United States of America **102**, 3459–3464.

**Cheng Y, Dai X, Zhao Y**. 2006. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. Genes and Development **20**, 1790–1799.

**Cheng Y, Dai X, Zhao Y**. 2007. Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. Plant Cell **19**, 2430–2439.

**Cheng J, Zhang M, Tan B, *et al.*** 2019. A single nucleotide mutation in *GID1c* disrupts its interaction with DELLA1 and causes a GA-insensitive dwarf phenotype in peach. Plant Biotechnology Journal **17**, 1723–1735.

**Cheval C, Aldon D, Galaud JP, Ranty B**. 2013. Calcium/calmodulin-mediated regulation of plant immunity. Biochimica et Biophysica Acta **1833**, 1766–1771.

**Choudhury S, Panda P, Sahoo L, Panda SK**. 2013. Reactive oxygen species signaling in plants under abiotic stress. Plant Signaling and Behavior **8**, e23681.

**Conesa A, Götz S**. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. International journal of plant genomics, 1–13.

**Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M**. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**, 3674–3676.

**Conesa A, Madrigal P, Tarazona S, *et al.*** 2016. A survey of best practices for RNA-seq data analysis. Genome Biology **17**, 1–19.

**Costa-Silva J, Domingues D, Lopes FM**. 2017. RNA-Seq differential expression analysis: An extended review and a software tool. PLOS ONE **12**, 1–18.

**Cummins I, Dixon DP, Freitag-Pohl S, Skipsey M, Edwards R**. 2011. Multiple roles for plant

glutathione transferases in xenobiotic detoxification. Drug Metabolism Reviews **43**, 266–280.

**Czyzewicz N, Nikonorova N, Meyer MR, Sandal P, Shah S, Vu LD, Gevaert K, Rao AG, De Smet I**. 2016. The growing story of (*Arabidopsis*) CRINKLY 4. Journal of Experimental Botany **67**, 4835–4847.

**Daccord N, Celton JM, Linsmith G, *et al.*** 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics **49**, 1099–1106.

**Dai X, Mashiguchi K, Chen Q, Kasahara H, Kamiya Y, Ojha S, DuBois J, Ballou D, Zhao Y**. 2013. The biochemical mechanism of auxin biosynthesis by an *Arabidopsis* YUCCA flavin-containing monooxygenase. Journal of Biological Chemistry **288**, 1448–1457.

**Daviere J-M, Achard P**. 2013. Gibberellin signaling in plants. Development **140**, 1147–1151.

**Delude C, Fouillen L, Bhar P, Cardinal MJ, Pascal S, Santos P, Kosma DK, Joubès J, Rowland O, Domergue F**. 2016. Primary fatty alcohols are major components of suberized root tissues of *Arabidopsis* in the form of Alkyl hydroxycinnamates. Plant Physiology **171**, 1934–1950.

**Deng Z, Zhang X, Tang W, *et al.*** 2007. A proteomics study of brassinosteroid response in *Arabidopsis*. Molecular and Cellular Proteomics **6**, 2058–2071.

**Dixon DP, Edwards R**. 2010. Glutathione transferases. *Arabidopsis* Book **8**, e0131.

**Dixon DP, Hawkins T, Hussey PJ, Edwards R**. 2009. Enzyme activities and subcellular localization of members of the *Arabidopsis* glutathione transferase superfamily. Journal of Experimental Botany.1207–1218.

**Dodds PN, Rathjen JP**. 2010. Plant immunity: Towards an integrated view of plant-pathogen interactions. Nature Reviews Genetics **11**, 539–548.

**Dong J, Huang H**. 2018. Auxin polar transport flanking incipient primordium initiates leaf adaxial-abaxial polarity patterning. Journal of Integrative Plant Biology **60**, 455–464.

**van Dyk MM, Soeker MK, Labuschagne IF, Rees DJG**. 2010. Identification of a major QTL for time of initial vegetative budbreak in apple (*Malus* x *domestica* Borkh.). Tree Genetics and Genomes **6**, 489–502.

**Eaton DL, Bammler TK**. 1999. Concise review of the glutathione S-transferases and their significance to toxicology. Toxicological Sciences **49**, 156–164.

**Eichmann R, Schäfer P**. 2015. Growth versus immunity-a redirection of the cell cycle? Current Opinion in Plant Biology **26**, 106–112.

**Elias AA, Busov VB, Kosola KR, Ma C, Etherington E, Shevchenko O, Gandhi H, Pearce DW, Rood SB, Strauss SH**. 2012. Green revolution trees: Semidwarfism transgenes modify gibberellins, promote root growth, enhance morphological diversity, and reduce competitiveness in hybrid *Poplar*. Plant Physiology **160**, 1130–1144.

**Espelie KE, Franceschi VR, Kolattukudy PE**. 1986. Immunocytochemical localization and time

course of appearance of an anionic peroxidase associated with suberization in wound-healing potato tuber tissue. Plant Physiology **81**, 487–492.

**Foyer CH, Noctor G**. 2009. Redox Regulation in Photosynthetic Organisms. Regulation **11**, 861–905.

**Francoz E, Ranocha P, Nguyen-Kim H, Jamet E, Burlat V, Dunand C**. 2015. Roles of cell wall peroxidases in plant development. Phytochemistry **112**, 15–21.

**Franke R, Briesen I, Wojciechowski T, Faust A, Yephremov A, Nawrath C, Schreiber L**. 2005. Apoplastic polyesters in *Arabidopsis* surface tissues - A typical suberin and a particular cutin. Phytochemistry **66**, 2643–2658.

**Franke R, Schreiber L**. 2007. Suberin - a biopolyester forming apoplastic plant interfaces. Current Opinion in Plant Biology **10**, 252–259.

**Galon Y, Finkler A, Fromm H**. 2010. Calcium-regulated transcription in plants. Molecular Plant **3**, 653–669.

**Gao QM, Zhu S, Kachroo P, Kachroo A**. 2015. Signal regulators of systemic acquired resistance. Frontiers in Plant Science **6**, 228.

**Ghosh A, Servin JA, Park G, Borkovich KA**. 2014. Global analysis of serine/threonine and tyrosine protein phosphatase catalytic subunit genes in Neurospora crassa reveals interplay between phosphatases and the p38 mitogen-activated protein kinase. G3: Genes, Genomes, Genetics **4**, 349–365.

**Gifford ML, Dean S, Ingram GC**. 2003. The *Arabidopsis* ACR4 gene plays a role in cell layer organisation during ovule integument and sepal margin development. Development **130**, 4249–4258.

**Gifford ML, Robertson FC, Soares DC, Ingram GC**. 2005. *Arabidopsis* CRINKLY4 function, internalization, and turnover are dependent on the extracellular crinkly repeat domain. Plant Cell **17**, 1154–1166.

**Gleason C, Huang S, Thatcher LF, Foley RC, Anderson CR, Carroll AJ, Millar AH, Singh KB**. 2011. Mitochondrial complex II has a key role in mitochondrial-derived reactive oxygen species influence on plant stress gene regulation and defense. Proceedings of the National Academy of Sciences of the United States of America **108**, 10768–10773.

**Götz S, Arnold R, Sebastian-Leon P, Martin-Rodriguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T, Conesa A**. 2011. B2G-FAR, a species-centered GO annotation repository. Bioinformatics **27**, 919–924.

**Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A**. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research **36**, 3420–3435.

**Griffiths J, Murase K, Rieu I, *et al.*** 2006. Genetic characterization and functional analysis of the GID1 gibberellin receptors in *Arabidopsis*. The Plant Cell **18**, 3399–3414.

**Gruner K, Griebel T, Návarová H, Attaran E, Zeier J**. 2013. Reprogramming of plants during

systemic acquired resistance. Frontiers in Plant Science **4**.

**Gullner G, Komives T, Király L, Schröder P**. 2018. Glutathione S-transferase enzymes in plant-pathogen interactions. Frontiers in Plant Science **871**, 1–19.

**Hammerschmidt R**. 1999. Induced disease resistance: How do induced plants stop pathogens? Physiological and Molecular Plant Pathology **55**, 77–84.

**Hattori A, Migitaka H, Ego M, Itoh M, Yamamoto K, Hara M, Suzuki T, Biology S, Antonio S**. 1995. Identification of melatonin in plants and its effects on plasma. Biochemistry and Molecular Biology International **35**, 627–634.

**Hedden P**. 2003. The genes of the Green Revolution. Trends in Genetics **19**, 5–9.

**Hedden P, Phillips AL**. 2000. Gibberellin metabolism: New insights revealed by the genes. Trends in Plant Science **5**, 523–530.

**Hernández-ruiz J**. 2006. The physiological function of melatonin in plants. Plant Signal and Behavior **1**, 89–95.

**Hiraga S, Sasaki K, Ito H, Ohashi Y, Matsui H**. 2001. A large family of class III plant peroxidases. Plant and Cell Physiology **42**, 462–468.

**Huang J, Tang D, Shen Y, *et al.*** 2010. Activation of gibberellin 2-oxidase 6 decreases active gibberellin levels and creates a dominant semi-dwarf phenotype in rice (*Oryza sativa* L.). Journal of Genetics and Genomics **37**, 23–36.

**Huerta-Cepas J, Szklarczyk D, Heller D, *et al.*** 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research **47**, D309–D314.

**Huot B, Yao J, Montgomery BL, He SY**. 2014. Growth-defense tradeoffs in plants: A balancing act to optimize fitness. Molecular Plant **7**, 1267–1287.

**Jeuken MJW, Zhang NW, McHale LK, Pelgrom K, Den Boer E, Lindhout P, Michelmore RW, Visser RGF, Niks RE**. 2009. Rin4 causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. Plant Cell **21**, 3368–3378.

**Jin P, Guo T, Becraft PW**. 2000. The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. Genesis **27**, 104–116.

**Jones J, Dangl J**. 2006. The plant immune system. Nature **444**, 323–329.

**Juvvadi P, Lamoth F, Steibach W**. 2014. Calcineurin-mediated regulation of hyphal growth, septation, and virulence in *Aspergillus fumigatus*. Mycopathologia **178**, 341–348.

**Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tsichlis PN, Makris AM**. 2000. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. Journal of Biological Chemistry **275**, 29207–29216.

**Kanehisa M, Goto S**. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research **28**, 27–30.

**Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M**. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research **44**, D457–D462.

**Kärkönen A, Kuchitsu K**. 2015. Reactive oxygen species in cell wall metabolism and development in plants. Phytochemistry **112**, 22–32.

**Karuppanapandian T, Moon JC, Kim C, Manoharan K, Kim W**. 2011. Reactive oxygen species in plants: Their generation, signal transduction, and scavenging mechanisms. Australian Journal of Crop Science **5**, 709–725.

**Kaurilind E, Xu E, Brosché M**. 2015. A genetic framework for $H_2O_2$ induced cell death in *Arabidopsis* thaliana. BMC Genomics **16**, 1–17.

**Khush GS**. 2001. Green revolution: the way forward. Nature Reviews Genetics **2**, 815–822.

**Komorisono M**. 2005. Analysis of the rice mutant dwarf and gladius leaf 1. aberrant Katanin-mediated microtubule organization causes up-regulation of gibberellin biosynthetic genes independently of gibberellin signaling. Plant Physiology **138**, 1982–1993.

**Krost C, Petersen R, Schmidt ER**. 2012. The transcriptomes of columnar and standard type apple trees (*Malus* x *domestica*) — A comparative study. Gene **498**, 223–230.

**Krüger J, Thomas CM, Golstein C, Dixon MS, Tang S, Mulder L, Jones JDG, Robinson E, Comp J**. 2002. A tomato cysteine protease required for *Cf(*-2-dependent disease resistance and suppression of autonecrosis. Science **296**, 744–747.

**Kuźniak E, Urbanek H**. 2000. The involvement of hydrogen peroxide in plant responses to stresses. Acta Physiologiae Plantarum **22**, 195–203.

**Labuschagné IF, Louw JH, Schmidt K, Sadie A**. 2002. Genetic variation in chilling requirement in apple progeny. Journal of the American Society for Horticultural Science **127**, 663–672.

**Lagrimini LM, Joly RJ, Dunlap JR, Liu TTY**. 1997. The consequence of peroxidase overexpression in transgenic plants on root growth and development. Plant Molecular Biology **33**, 887–895.

**Langmead**. 2013. Fast gapped-read alignment with Bowtie 2. Nature Methods **9**, 357–359.

**Legave J-M, Guédon Y, Malagi G, El Yaacoubi A, Bonhomme M**. 2015. Differentiated responses of apple tree floral phenology to global warming in contrasting climatic regions. Frontiers in Plant Science **6**, 1053.

**Leubner-metzger G**. 2011. Members of the gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. Journal of Experimental Botany, **62**, 5131–5147.

**Li L, Yu Y, Zhou Z, Zhou J**. 2016. Plant pattern-recognition receptors controlling innate immunity. **59**, 878–888.

**Liang X, Zhang L, Natarajan SK, Becker DF**. 2013. Proline mechanisms of stress survival. Antioxidants and Redox Signaling **19**, 998–1011.

**Liu N, Sun Y, Pei Y, Zhang X, Wang P, Li X, Li F, Hou Y**. 2018. A pectin methylesterase inhibitor enhances resistance to verticillium wilt. Plant Physiology **176**, 2202–2220.

**Lotfy S, Javelle F**. 1994. Formation of ω-feruloyloxypalmitic acid by an enzyme from wound-healing potato tuber discs. Phytochemistry **35**, 1419–1424.

**Lotfy S, Javelle F, Negrel J**. 1996. Purification and characterization of hydroxycinnamoyl- Coenzyme A: ω- hydroxypalmitic acid O-hydroxycinnamoyltransferase from tobacco (*Nicotiana tabacum* L.) cell-suspension cultures. Planta **199**, 475–480.

**Love MI, Huber W, Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**, 550.

**Macho AP, Zipfel C**. 2014. Plant PRRs and the activation of innate immune signaling. Molecular Cell **54**, 263–272.

**Marrs KA**. 1996. The functions and regulation of glutathione s-transferases in plants. Annual Review of Plant Physiology and Plant Molecular Biology **47**, 127–158.

**Martin DN, Proebsting WM, Hedden P**. 1999. The *SLENDER* gene of pea encodes a gibberellin 2-oxidase. Plant Physiology **121**, 775–781.

**Melke A, Fetene M**. 2014. Apples (*Malus domestica*, Borkh.) phenology in Ethiopian highlands: Plant growth, blooming, fruit development and fruit quality perspectives. American Journal of Experimental Agriculture **4**, 1958–1995.

**Mittler R**. 2002. Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science **7**, 405–410.

**Mittler R, Vanderauwera S, Gollery M, Breusegem F Van**. 2004. Reactive oxygen gene network of plants. Trends in Plant Science **9**, 490–498.

**Mohan R, Bajar A, Kolattukudy P**. 1993*a*. Induction of a tomato anionic peroxidase gene (*tap1*) by wounding in transgenic tobacco and activation of *tap1*/*GUS* and *tap2*/*GUS* chimeric gene fusions in transgenic tobacco by wounding and pathogen attack. Plant molecular biology **21**, 341–354.

**Mohan R, Vijayan P, Kolattukudy P**. 1993*b*. Developmental and tissue-specific expression of a tomato anionic peroxidase (*tap1*) gene by a minimal promoter, with wound and pathogen induction by an additional 5′-flanking region. Plant molecular biology **22**, 475–490.

**Molina I, Li-Beisson Y, Beisson F, Ohlrogge JB, Pollard M**. 2009. Identification of an *Arabidopsis* feruloyl-coenzyme a transferase required for suberin synthesis. Plant Physiology **151**, 1317–1328.

**Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B**. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods **5**, 621.

**Naoumkina MA, Zhao Q, Gallego-Giraldo L, Dai X, Zhao PX, Dixon RA**. 2010. Genome-wide

analysis of phenylpropanoid defence pathways. Molecular Plant Pathology **11**, 829–846.

**Nygren PJ, Scott JD**. 2016. Regulation of the phosphatase PP2B by protein-protein interactions. Biochemical Society Transactions **44**, 1313–1319.

**O'Brien JA, Daudi A, Butt VS, Bolwell GP**. 2012. Reactive oxygen species and their role in plant defence and cell wall metabolism. Planta **236**, 765–779.

**Olszewski N, Sun T, Gubler F**. 2002. Gibberellin signaling: biosynthesis, catabolism, and response pathways. The Plant cell **14 (Suppl.**, S61-80.

**Orr H**. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics **144**, 1331–1335.

**Orr H, Presgraves D**. 2000. Speciation by postzygotic isolation: forces, genes and molecules. BioEssays **22**, 1085–1094.

**Otani M, Meguro S, Gondaira H, *et al.*** 2013. Overexpression of the gibberellin 2-oxidase gene from Torenia fournieri induces dwarf phenotypes in the liliaceous monocotyledon *Tricyrtis* sp. Journal of Plant Physiology **170**, 1416–1423.

**Ou C, Jiang S, Wang F, Tang C, Hao N**. 2015. An RNA-Seq analysis of the pear (*Pyrus communis* L.) transcriptome , with a focus on genes associated with dwarf. Plant Gene **4**, 69–77.

**Passardi F, Cosio C, Penel C, Dunand C**. 2005. Peroxidases have more functions than a Swiss army knife. Plant Cell Reports **24**, 255–265.

**Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP**. 1997. The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. , 3194–3205.

**Perl-Treves R, Foley RC, Chen W, Singh KB**. 2004. Early induction of the *Arabidopsis* GSTF8 promoter by specific strains of the fungal pathogen Rhizoctonia solani. Molecular Plant-Microbe Interactions **17**, 70–80.

**Pieterse C, van der Does D, Zamioudis C, Leon-Reyes A, van Wees S**. 2012. Hormonal modulation of plant immunity. Annual Review of Ce **28**, 489–521.

**Poethig RS**. 2010. The past, present, and future of vegetative phase change. Plant Physiology **154**, 541–544.

**Poovaiah BW, Du L, Wang H, Yang T**. 2013. Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. Plant Physiology **163**, 531–542.

**Porto DD, Bruneau M, Perini P, Anzanello R, Renou J**. 2015. Transcription profiling of the chilling requirement for bud break in apples : a putative role for FLC-like genes.

**Pusztahelyi T**. 2018. Chitin and chitin-related compounds in plant–fungal interactions. Mycology **9**, 189–201.

**Qi J, Wang Y, Yu T, Cunha A, Wu B, Vernoux T, Meyerowitz E, Jiao Y**. 2014. Auxin depletion

from leaf primordia contributes to organ patterning. Proceedings of the National Academy of Sciences of the United States of America **111**, 18769–18774.

**R Core Team**. 2018. R *:* A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

**Ranathunge K, Schreiber L, Franke R**. 2011. Suberin research in the genomics era — New interest for an old polymer. Plant Science **180**, 399–413.

**Reinhardt D, Mandel T, Kuhlemeier C**. 2000. Auxin regulates the initiation and radial position of plant lateral organs. Plant Cell **12**, 507–518.

**Rieseberg LH, Blackman BK**. 2010. Speciation genes in plants. Annals of Botany **106**, 439–455.

**Rieseberg LH, Willis JH**. 2007. Plant speciation. Science **317**, 910–914.

**Rieu I, Eriksson S, Powers SJ, *et al.*** 2008. Genetic analysis reveals that C 19-GA 2-oxidation is a major gibberellin inactivation pathway in *Arabidopsis*. The Plant Cell **20**, 2420–2436.

**Robert-Seilaniantz A, Grant M, Jones J**. 2010. Hormone crosstalk in plant disease and defense: More than just jasmonate-salicylate antagonism. Annual Review of Phytopathologyhytopathology **49**, 317–343.

**Sappl PG, Oñate-Sánchez L, Singh KB, Millar AH**. 2004. Proteomic analysis of glutathione S-transferases of *Arabidopsis* thaliana reveals differential salicylic acid-induced expression of the plant-specific phi and tau classes. Plant Molecular Biology **54**, 205–219.

**Scandalios JG**. 2005. Oxidative stress: Molecular perception and transduction of signals triggering antioxidant gene defenses. Brazilian Journal of Medical and Biological Research **38**, 995–1014.

**Schomburg FM, Bizzell CM, Lee DJ, Zeevaart JAD, Amasino RM**. 2003. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. The Plant Cell **15**, 151–163.

**Serra O, Hohn C, Franke R, Prat S, Molinas M, Figueras M**. 2010. A feruloyl transferase involved in the biosynthesis of suberin and suberin-associated wax is required for maturation and sealing properties of potato periderm. Plant Journal **62**, 277–290.

**Shi HY, Li ZH, Zhang YX, Chen L, Xiang DY, Zhang YF**. 2014. Two pear glutathione S-transferases genes are regulated during fruit development and involved in response to salicylic acid, auxin, and glucose signaling. PLoS One **9**, e89926.

**Shi H, Wei Y, Wang Q, Reiter RJ, He C**. 2016. Melatonin mediates the stabilization of DELLA proteins to repress the floral transition in *Arabidopsis*. Journal of Pineal Research **60**, 373–379.

**Shigeto J, Tsutsumi Y**. 2016. Diverse functions and reactions of class III peroxidases. New Phytologist **209**, 1395–1402.

**Singh DP, Jermakow AM, Swain SM**. 2002. Gibberellins are required for seed development and pollen tube growth in *Arabidopsis*. The Plant Cell **14**, 3133–3147.

**Solfanelli C, Ceron F, Paolicchi F, Giorgetti L, Geri C, Ceccarelli N, Kamiya Y, Picciarelli P**. 2005. Expression of two genes encoding gibberellin 2- and 3-oxidases in developing seeds of *Phaseolus coccineus*. Plant and Cell Physiology **46**, 1116–1124.

**Sun T**. 2008. Gibberellin metabolism, perception and signaling pathways in *Arabidopsis*. The *Arabidopsis* Book **6**, e0103.

**Sun T**. 2011. The Molecular mechanism and evolution of the GA–GID1–DELLA signaling module in plants review. Current Biology **21**, R338–R345.

**Sun T, Gubler F**. 2004. Molecular mechanism of gibberellin signalling in plants. Annual Review of Plant Biology **55**, 197–223.

**Takenaka Y, Nakano S, Tamoi M, Sakuda S, Fukamizo T**. 2009. Chitinase gene expression in response to environmental stresses in *Arabidopsis* thaliana: Chitinase inhibitor allosamidin enhances stress tolerance. Bioscience, Biotechnology and Biochemistry **73**, 1066–1071.

**Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y**. 2001. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopis embryos and juvenile plants. Development **128**, 4681–4689.

**Tanaka H, Watanabe M, Sasabe M, Hiroe T, Tanaka T, Tsukaya H, Ikezaki M, Machida C, Machida Y**. 2007. Novel receptor-like kinase ALE2 controls shoot development by specifying epidermis in *Arabidopsis*. Development **134**, 1643–1652.

**Tanaka H, Watanabe M, Watanabe D, Tanaka T, Machida C, Machida Y**. 2002. ACR4, a putative receptor kinase gene of *Arabidopsis* thaliana, that is expressed in the outer cell layers of embryos and plants, is involved in proper embryogenesis. Plant and Cell Physiology **43**, 419–428.

**Tatusov RL, Galperin MY, Natale DA, Koonin EV**. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research **28**, 33–36.

**Thaler JS, Humphrey PT, Whiteman NK**. 2012. Evolution of jasmonate and salicylate signal crosstalk. Trends in Plant Science **17**, 260–270.

**Thatcher LF, Carrie C, Andersson CR, Sivasithamparam K, Whelan J, Singh KB**. 2007. Differential gene expression and subcellular targeting of *Arabidopsis* glutathione S-transferase F8 is achieved through alternative transcription start sites. Journal of Biological Chemistry **282**, 28915–28928.

**Thatcher LF, Kamphuis LG, Hane JK, Oñate-Sánchez L, Singh KB**. 2015. The *Arabidopsis* KH-domain RNA-binding protein ESR1 functions in components of jasmonate signalling, unlinking growth restraint and resistance to stress. PLoS One **10**, 1–31.

**Thomas SG, Phillips AL, Hedden P**. 1999. Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. Proceedings of the National Academy of Sciences **96**, 4698–4703.

**Thomma BPHJ, Nürnberger T, Joosten MHAJ**. 2011. Of PAMPs and effectors: The blurred PTI-

ETI dichotomy. Plant Cell **23**, 4–15.

**Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L**. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology **28**, 511–515.

**Uhrig RG, Labandera AM, Moorhead GB**. 2013. *Arabidopsis* PPP family of serine/threonine protein phosphatases: Many targets but few engines. Trends in Plant Science **18**, 505–513.

**Velasco R, Zharkikh A, Affourtit J, *et al.*** 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genetics **42**, 833–839.

**Vishwanath SJ, Kosma DK, Pulsifer IP, Scandola S, Pascal S, Joubès J, Dittrich-Domergue F, Lessire R, Rowland O, Domergue F**. 2013. Suberin-associated fatty alcohols in *Arabidopsis*: Distributions in roots and contributions to seed coat barrier properties. Plant Physiology **163**, 1118–1132.

**Vleesschauwer D De, Xu J, Höfte M**. 2014. Making sense of hormone-mediated defense networking : from rice to *Arabidopsis*. **5**, 1–15.

**Wagner U, Edwards R, Dixon DP, Mauch F**. 2002. Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. Plant Molecular Biology **49**, 515–532.

**Wang Q, An B, Shi H, Luo H, He C**. 2017. High Concentration of Melatonin Regulates Leaf Development by Suppressing Cell Proliferation and Endoreduplication in *Arabidopsis*.

**Wang Q, An B, Wei Y, Reiter RJ, Shi H, Luo H, He C**. 2016. Melatonin regulates root meristem by repressing auxin synthesis and polar auxin transport in *Arabidopsis*. Frontiers in Plant Science **7**, 1882.

**Wang Z, Gerstein M, Snyder M**. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics **10**, 57–63.

**Wang Y, Reiter RJ, Chan Z**. 2018. Phytomelatonin: A universal abiotic stress regulator. Journal of Experimental Botany **69**, 963–974.

**Wang W, Xu B, Wang H, Li J, Huang H, Xu L**. 2011. *YUCCA* genes are expressed in response to leaf adaxial-abaxial juxtaposition and are required for leaf margin development. Plant Physiology **157**, 1809–1819.

**Watanabe M, Tanaka H, Watanabe D, Machida C, Machida Y**. 2004. The ACR4 receptor-like kinase is required for surface formation of epidermis-related tissues in *Arabidopsis* thaliana. Plant Journal **39**, 298–308.

**Xu H, Cao D, Chen Y, Wei D, Wang Y, Stevenson RA, Zhu Y, Lin J**. 2016. Gene expression and proteomic analysis of shoot apical meristem transition from dormancy to activation in *Cunninghamia lanceolata* (Lamb.) Hook. Scientific Reports **6**, 19938.

**Zhang L, Becker DF**. 2015. Connecting proline metabolism and signaling pathways in plant senescence. Frontiers in Plant Science **6**, 1–8.

**Zhang C-C, Wang L-Y, Wei K, Wu L-Y, Li H-L, Zhang F, Cheng H, Ni D-J**. 2016. Transcriptome analysis reveals self-incompatibility in the tea plant (*Camellia sinensis*) might be under gametophytic control. BMC Genomics **17**, 359.

**Zhang Z, Zhuo X, Zhao K, Zheng T, Han Y, Yuan C, Zhang Q**. 2018. Transcriptome profiles reveal the crucial roles of hormone and sugar in the bud dormancy of *Prunus mume*. Scientific Reports **8**, 5090.

# Chapter 6
# General Discussions and Future Prospects

## 6.1 Project overview and findings

Dwarf traits such as crinkle dwarf have not been explored extensively in apple, but is considered economically unfavourable due to their undesirable characteristics for production. Crinkle dwarf phenotype can be considered as a type of hybrid incompatibility, in this case hybrid necrosis, where progeny show symptoms of permanent activation of stress responses, such as dwarfism and tissue necrosis. It will be useful to eliminate a typical seedling phenotype associated with crinkled leaves, poor growth, and in some cases lethality, from apple breeding lines. Therefore understanding the genes

involved in controlling these phenotypes and identifying molecular markers linked to the trait are of interest to the ARC and other similar breeding programmes.

In this research, the unveiling of the molecular genetics underlying the crinkle dwarf trait was pursued by addressing three main objectives: 1) to generate relevant crosses with the aim to investigate the inheritance patterns of the crinkle dwarf trait, 2) to elucidate inheritance patterns by molecular mapping using the apple 20K Infinium® single nucleotide polymorphisms (SNP) array (Bianco *et al.*, 2014), and, 3) to identify genes that are differentially expressed in normal *versus* crinkle dwarf phenotypes using RNA-sequencing technologies.

***Inheritance patterns of crinkle dwarf phenotype***

Thirteen F1 apple progenies were raised to investigate the inheritance of the crinkle dwarf trait, initially by studying their segregation patterns. The phenotypic evaluation of segregating population of the primary cross 'McIntosh' x Malling 9 ('M.9'), both parents of normal habit, segregated 9:7 for normal *versus* crinkle dwarf, which is consistent with a two-gene control and for which the two parents are heterozygous (*DdEe*). The crinkle dwarf phenotype occurs when either of the two genes are homozygous recessive *D-ee* or *ddE-*, suggesting two epistatic genes segregating simultaneously. An additional twelve progenies were raised with the aim to target progenies that segregate for one of the two loci. Seven progenies segregated 3:1, with possible parental genotypes e.g. *DdEe* x *DDEe,* suggesting control by one gene for which crinkle dwarf phenotype would be expressed when one gene is homozygous recessive (*DDee*, *Ddee*).

The success of any breeding system is governed by the optimum parental cross combinations and the direction of the cross. Bošković and Tobutt, (1999) reported that some apple cultivars are incompatible when crossed, and some triploid x diploid combinations fail whereas their reciprocals succeed. In the present study, the direction of the cross and the parental self-incompatibility genotypes were not considered prior to experimental design of cross combinations, and in turn resulted in some progenies not segregating for the crinkle dwarf phenotype. The apparent lack of segregation suggested a possibility of crinkle dwarf being *S*-linked, however this was rejected based on the mapping results

from chapter four. The findings of the study suggest that crinkle dwarf phenotype may rather be a case of hybrid incompatibility resulting in and linked to distortion segregation.

Self-incompatibility is an important genetic mechanism that prevents inbreeding and promotes genetic heterogeneity in flowering plants (DeNettancourt, 1997; McClure *et al.*, 2011). Traditionally, self-incompatibility interactions are determined by performing controlled crosses and evaluating pollen tube growth through the style, through microscopy (Bošković and Tobutt, 1996; DeNettancourt, 1997; Herrera *et al.*, 2018). However, information regarding self-incompatibility has since evolved over the years and molecular tools based on the sequences of the *S*-locus, such as PCR-based consensus and allele-specific primers have been developed and characterised in most fruit crops, including apple (Sakurai *et al.*, 2000; Broothaerts, 2003; De Franceschi *et al.*, 2016). In this study, nine *S*-genotypes were determined and 'Irish Peach' ($S_1S_{37}$) and 'Howgate Wonder' ($S_3S_5$) confirmed herein. The S-genotypes for the rootstock cultivar Malling 1 ('M.1') ($S_3S_9$) and the selection TSR1T187 ($S_7S_{24}$) were deduced for the first time.

To reduce the challenges posed by self-incompatibility, knowledge of cross-compatibility is important. Furthermore, it is used to select optimum cross combinations based on known S-genotypes which enables predicted outcomes, which in turn, maximizes breeding efficiency and fruit quality (Orcheski and Brown, 2012). Moreover, the knowledge of S-genotypes is also important in the establishment of trueness-to-parentage.

***Molecular mapping***

Recent advances in high throughput sequencing technologies have created the opportunity to obtain genome-wide genetic markers for high-resolution linkage mapping through the use of SNP chip arrays, in a more time- and cost-effective manner (Bianco *et al.*, 2014; Nadeem *et al.*, 2018). High density linkage maps are important tools in both traditional and more modern fruit breeding programmes as they constitute the framework in studying the co-inheritance of markers and traits, and in turn facilitate marker-assisted selection (MAS) (Troggio *et al.*, 2007; Nadeem *et al.*, 2018).

For mapping the crinkle dwarf phenotype, a single progeny of 'McIntosh' x 'M.1' consisting of 118 seedlings and segregating 90:28 ($\chi^2$ = 0.10, ρ=0.750), a 3:1 ratio, presumably only for one of the two loci (*D-ee* or *ddE-*) was selected for mapping. High-density genetic maps were constructed for the two parents using the apple 20K Infinium® SNP array. The crinkle dwarf trait mapped to linkage group (LG) 8 in 'McIntosh' at position 54.38 cM, but unexpectedly, to LG2 in 'M.1' at position 5.63 cM. It was noted that LG8 of 'M.1' was comparatively shorter than that of 'McIntosh' with a total length of 45.13 cM, and lacked the target region where crinkle dwarf mapped on 'McIntosh' (Figure 4.3). Therefore, LG8 of the 'M.1' genetic map was compared against the physical map (Figure 4.4). It was confirmed that LG8 was well covered with markers, however with monomorphic SNPs at the target region. Poorly saturated genomic regions does not necessarily reflect a shortage of SNP markers, but rather a lack of polymorphic and informative ones (van Berloo *et al.*, 2008; Antanaviciute *et al.*, 2012).

A similar trend of poor coverage due to monomorphic SNPs on target regions have been reported in tomatoes for example (van Berloo *et al.*, 2008). Therefore, the lack of the target region in 'M.1', might have resulted in an incorrect mapping of crinkle dwarf trait on LG2 in 'M.1', and this has to be investigated further. Consequently, the LG8 of both 'McIntosh' and 'M.1' are currently being validated with published apple (*Malus*) microsatellite (SSR) to further dissect the target region.

Additionally, a significance test for validation of the nearest SNP marker associated with the crinkle dwarf trait was performed in 'McIntosh'. The highest Kruskal-Wallis (KW) score of 30.67 ($p < 0.0001$) declared a significant marker-trait association on LG8 which co-segregated with marker SNP_FB_0765586 at a genetic position of 54.28 cM. Validation of the findings is however needed and should be fine-mapped in other populations segregating for crinkle dwarf trait with larger number of seedlings. Moreover, crinkle dwarf trait also mapped on LG8 on the consensus genetic linkage map, however co-segregated with marker (GDsnp02575) (Figure 4.6). Segregation distortion was observed in both the parental genetic maps and consensus genetic map with varying degree of deviation. These findings also suggest that the crinkle dwarf locus may be situated close to lethal genes.

Moreover, a user-friendly marker/probe, preferably SSR, that is specific for crinkle dwarf trait will in future be designed at the ARC, and utilised in screening other mapping progenies segregating for the

trait. The use of the probe will enable the elimination of crinkle dwarf growth habit by screening the breeding lines at an early seedling stage. In the case of crinkle dwarfs, marker assisted selection (MAS) can be employed as early as one week after germination, since the actual phenotype manifest distinctly 12 weeks after germination.

***Transcriptome Profiling***

Currently, the molecular mechanisms underlying crinkle dwarf trait in apple, is poorly understood. In the present study, a transcriptome profile was generated from pooled tissues (apical buds and young leaves) collected at different developmental stages between the contrasting phenotypes using RNA sequencing technology. A total of 921 (763 up-regulated and 158 down-regulated) transcripts were significantly, differentially expressed at log2foldchange > +1.5 and < -1.5 at p-adjusted < 0.05, and a false discovery rate (FDR) of $p < 0.01$ were obtained. The major biological categories of the DEGs obtained from KEGG biological pathway analysis corresponded to those involved in defense responses, oxidative stress, pathogenesis-related genes, secary metabolite and cell wall modification.

One of the most important events in the early phase of incompatible plant–pathogen interaction is the rapid and transient production of reactive oxygen species (ROS). The accumulation of ROS is involved in regulating cell death (Fujioka and Yokota, 2003; Schmidt and Schippers, 2015). In this regard, growing evidence shows that the generation of ROS is one of the most common plant responses to different stresses (Sewelam *et al.*, 2016). Other studies have suggested that oxidative stress plays a role in fitness trade-offs in life-history evolution and functional ecology (Costantini *et al.*, 2010; Latta *et al.*, 2019). Barreto and Burton (2013), stated that, the extent of oxidative stress in hybrids appears to be dependent on the degree of genetic divergence between their respective parental populations. In the present study, lactoperoxidase (E.C. 1.11.1.7) (Prx), was the most highly expressed enzyme along with the activation of antioxidant enzyme, glutathione *S*-transferase (GST). Both Prx and GST are a source of ROS (Passardi *et al.*, 2005; Gullner *et al.*, 2018). Therefore, generation of ROS has been implicated in various processes of programmed cell death (Sharma *et al.*, 2012). In the study of wheat necrosis, Prx was reported to increase with the progression of hybrid necrosis, which also suggested Prx to be a source of hydrogen peroxide. In the same study of wheat, antioxidant system was not well-coordinated

and resulted in oxidative stress which later caused cell death (Sharma *et al.*, 2003). Importantly, Sharma (2003), emphasized that the response of necrosis differed in magnitude at different developmental stages of leaves.

The present study also showed the activation of enzymes encoding for chitinase and pectin responsible for pattern recognition receptors, which are implicated in fungal pathogenesis attack. Along with, there were activation of genes encoding calcium dependent protein kinases (serine/threonine protein, in particular, calmodulin. Calcium signalling has been implicated in the hypersensitive reaction (HR) induced cell death using calcium channel blockers (Pennell and Lamb, 1997; Nygren and Scott, 2016). The molecular mechanisms for HR involves generation of oxidative burst in infected cells and programmed cell death (Mizuno *et al.*, 2010).

Phytohormones regulate almost every aspect of plant growth in response to developmental and environmental cues (Zheng *et al.*, 2015). They are also known to be regulated in response to wounding and insect attack. A signal hormonal network was observed with the up-regulation of alpha-linolenic, a precursor of jasmonic acid (JA) and suppression of growth associated hormone pathway, 2beta-dioxygenase (*GA2ox*). Jasmonic acid is induced by pathogen attack or wounding, which often leads to the generation of reactive oxygen species (ROS), including hydrogen peroxide (Creelman and Mullet, 1997; Savatin *et al.*, 2014).

Collectively, the findings of the present study coincide with the highlights proposed by Bomblies *et al*. (2007) and other hybrid incompatibility (hybrid necrosis) studies (Sharma *et al.*, 2003; Mizuno *et al.*, 2010; Świadek *et al.*, 2017), suggesting that crinkle dwarf phenotype may be an autoimmune type of response.

Finally, the scope of the transcriptome profiling is being extended and conducted in parallel with the current study. The same tissue extracts used in the present study are being investigated further on a larger scale, by sequencing cDNA libraries independently i.e. without an adoption of pooling strategies, with the aim to increase the coverage of the sequencing data and the statistical power of the differential

expression. Results obtained from the current and the newly conducted experiment will be merged and also subjected to real-time quantitative PCR (RT-qPCR) for validation.

In summary, the research presented in this dissertation have broadened the understanding of the molecular mechanisms underlying the crinkle dwarf phenotype. The knowledge of S-genotypes and MAS is key in reducing insufficient land usage, cost, and other breeding resources required to grow and maintain the plants that will later be discarded from the breeding lines. Further validation of the QTL is necessary and will be running in parallel.

## 6.2 References

**Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troggio M, Sargent DJ**. 2012. Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* infinium whole genome genotyping array. BMC Genomics **13**, 203.

**Bianco L, Cestaro A, Sargent DJ, *et al.*** 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh.). PLoS One **9**, e110377.

**Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D**. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biology **5**, 1962–1972.

**Bošković R, Tobutt KR**. 1996. Correlation of stylar ribonuclease zymograms with incompatibility alleles in sweet cherry. Euphytica **90**, 245–250.

**Broothaerts W**. 2003. New findings in apple *S*-genotype analysis resolve previous confusion and request the re-numbering of some *S*-alleles. Theoretical and Applied Genetics **106**, 703–714.

**Byrne DH**. 2012. Trends in fruit breeding. *In*: Badenes, ML and Byrne D, (Eds). *Fruit Breeding*. Boston, MA: Springer. pp. 3–36.

**DeNettancourt D**. 1997. Incompatibility in angiosperms. Sexual Plant Reproduction **10**, 185–199.

**De Franceschi P, Cova V, Tartarini S, Dondini L**. 2016. Characterization of a new apple S-RNase allele and its linkage with the *Rvi5* gene for scab resistance. Molecular Breeding **36**, 1–11.

**Herrera S, Lora J, Hormaza JI, Herrero M, Rodrigo J**. 2018. Optimizing production in the new generation of apricot cultivars: Self-incompatibility, S-RNase allele identification, and incompatibility group assignment. Frontiers in Plant Science **9**, 1–12.

**Hollender CA, Dardick C**. 2015. Molecular basis of angiosperm tree. New Phytologist **206**, 541–556.

**McClure B, Cruz-García F, Romero C**. 2011. Compatibility and incompatibility in S-RNase-based systems. Annals of Botany **108**, 647–658.

**McDermott SR, Noor MAF**. 2010. The role of meiotic drive in hybrid male sterility. Philosophical Transactions of the Royal Society B: Biological Sciences **365**, 1265–1272.

**Nadeem MA, Nawaz MA, Shahid MQ, *et al.*** 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment **32**, 261–285.

**Orcheski B, Brown S**. 2012. A grower's guide to self and cross-incompatibility in apple. New York Fruit Quarterly **20**, 25–28.

**Sakurai K, Brown SK, Weeden N**. 2000. Self-incompatibility alleles of apple cultivars and advanced selections. HortScience **35**, 116–119.

**Troggio M, Malacarne G, Coppola G, *et al.*** 2007. A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring pinot noir bacterial artificial chromosome contigs. Genetics **176**, 2637–2650.

**Van Berloo R, Zhu A, Ursem R, Verbakel H, Gort G, Van Eeuwijk FA**. 2008. Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theoretical and Applied Genetics **117**, 89–101.

# Supplementary Data

**Supplementary Table 3.1 Morphological characteristics of 'McIntosh' x 'M.1' measured between normal and crinkle dwarf seedlings 6 months after germination.**

| Seedling | Phenotype | seedling height (cm) | Diameter at b/h (mm) | Diameter at mid (mm) | Diameter at tip (mm) |
|---|---|---|---|---|---|
| **1** | crinkle | 8,5 | 1,78 | 1,78 | 1,19 |
| **2** | crinkle | 6,99 | 1,99 | 1,69 | 0,9 |
| **3** | crinkle | 14,9 | 3,2 | 2,4 | 3,46 |
| **4** | crinkle | 7,2 | 1,72 | 1,58 | 1,57 |
| **5** | crinkle | 7,3 | 1,81 | 2,61 | 2,52 |
| **6** | crinkle | 8,7 | 1,29 | 1,27 | 1,22 |
| **7** | crinkle | 12,1 | 1,46 | 1,55 | 0,95 |
| **8** | crinkle | 12,8 | 2,43 | 2,44 | 2,38 |
| **9** | crinkle | 15,2 | 1,87 | 1,62 | 1,33 |
| **10** | crinkle | 6 | 1,89 | 1,9 | 0,95 |
| **11** | crinkle | 8,7 | 1,34 | 1,3 | 1,6 |
| **12** | crinkle | 11,3 | 2,09 | 2,25 | 2,69 |
| **13** | crinkle | 18 | 3,25 | 2,55 | 2,72 |
| **14** | crinkle | 16,5 | 2,86 | 2,24 | 2,57 |
| **15** | crinkle | 13,9 | 2,8 | 2,21 | 2,38 |
| **16** | crinkle | 14,44 | 3,63 | 3,15 | 2,61 |
| **17** | crinkle | 15,32 | 4,75 | 4,85 | 3,54 |
| **18** | crinkle | 9,6 | 2,8 | 2,39 | 1,92 |
| **19** | crinkle | 8 | 1,44 | 2,04 | 2,55 |
| **20** | crinkle | 13,6 | 1,94 | 2,15 | 2,94 |
| **21** | crinkle | 9,2 | 1,5 | 1,6 | 1,36 |
| **22** | crinkle | 13 | 2,06 | 2,04 | 2,73 |
| **23** | crinkle | 7,99 | 1,57 | 1,67 | 1,59 |
| **24** | crinkle | 10,4 | 1,35 | 2,09 | 2,79 |
| **Average** | | 11,24 | 2,20 | 2,14 | 2,10 |
| | | | | | |
| | | | | | |

| Seedling | Phenotype | seedling height (cm) | Diameter at b/h (mm) | Diameter at mid (mm) | Diameter at tip (mm) |
|---|---|---|---|---|---|
| **25** | normal | 58,8 | 5,07 | 5,33 | 3,42 |
| **26** | normal | 114,5 | 8,63 | 4,7 | 2,82 |
| **27** | normal | 54 | 8,18 | 2,88 | 2,68 |
| **28** | normal | 117,5 | 9,12 | 4,17 | 2,51 |
| **30** | normal | 24,5 | 8,46 | 3,99 | 2,99 |
| **31** | normal | 92,5 | 8,13 | 5,65 | 3,62 |
| **32** | normal | 63,4 | 6,24 | 4,69 | 3,4 |
| **33** | normal | 99,5 | 5,17 | 4,75 | 2,96 |
| **34** | normal | 14,5 | 8,91 | 5,63 | 2,51 |
| **35** | normal | 73,4 | 7,51 | 4,45 | 2,14 |
| **36** | normal | 104,2 | 8,57 | 4,82 | 2,16 |
| **37** | normal | 67,1 | 8,33 | 5,12 | 4,1 |
| **38** | normal | 51,6 | 4,72 | 4,84 | 2,08 |
| **39** | normal | 85,2 | 10,08 | 4,03 | 3,3 |
| **40** | normal | 36,2 | 8,1 | 5,61 | 2,76 |
| **41** | normal | 85,8 | 9,47 | 6,76 | 2,84 |
| **42** | normal | 51,9 | 6,87 | 5,79 | 2,28 |
| **43** | normal | 88,3 | 7,87 | 5,12 | 2,41 |
| **44** | normal | 61,5 | 7,67 | 5,83 | 2,5 |
| **45** | normal | 68,9 | 6,8 | 4,23 | 2,48 |
| **46** | normal | 17 | 6,94 | 5,38 | 2,9 |
| **47** | normal | 96,5 | 5,83 | 5,19 | 2,67 |
| **48** | normal | 21,5 | 7,63 | 5,78 | 3,28 |
| **49** | normal | 86,2 | 7,14 | 4,51 | 2,85 |
| **50** | normal | 21,5 | 7,63 | 5,78 | 3,28 |
| **51** | normal | 55 | 7,72 | 4,69 | 2,53 |
| **52** | normal | 81,3 | 7,21 | 5,13 | 2,69 |
| **53** | normal | 87,2 | 7,71 | 5,22 | 3,57 |
| **54** | normal | 72,7 | 9,16 | 4,94 | 2,37 |
| **55** | normal | 11,4 | 6,46 | 5,49 | 2,12 |
| **56** | normal | 13,2 | 6,33 | * | * |
| **57** | normal | 98,5 | 8,34 | 5,01 | 2,62 |
| **58** | normal | 92,5 | 9,02 | 4,63 | 3,71 |
| **59** | normal | 56,1 | 5,53 | 3,65 | 3,21 |
| **60** | normal | 38,8 | 5,35 | 4,54 | 2,68 |

| | | | | | |
|---|---|---|---|---|---|
| **61** | normal | 21,6 | 3,01 | 2,69 | 2,12 |
| **62** | normal | 44 | 3,66 | 3,56 | 2,39 |
| **63** | normal | 33,1 | 3,58 | 3,28 | 2,57 |
| **64** | normal | 12,7 | 3,74 | 2,19 | 3,13 |
| **65** | normal | 63,4 | 5,49 | 3,76 | 2,52 |
| **66** | normal | 27,1 | 4,56 | 2,84 | 2,05 |
| **67** | normal | 20,1 | 3,56 | 2,4 | 2,13 |
| **68** | normal | 39,5 | 4,45 | 3,94 | 3,07 |
| **69** | normal | 84,1 | 6,19 | 4,31 | 3,4 |
| **70** | normal | 34,6 | 5,35 | 3,69 | 2,66 |
| **71** | normal | 21,3 | 3,06 | 2,74 | 3,06 |
| **72** | normal | 40,1 | 4,48 | 4,28 | 3,52 |
| **73** | normal | 98,5 | 5,98 | 4,12 | 2,48 |
| **74** | normal | 54,3 | 5,7 | 4,66 | 3,06 |
| **75** | normal | 35,8 | 5,34 | 4,09 | 2,75 |
| **76** | normal | 91,7 | 6,68 | 4,25 | 2,46 |
| **77** | normal | 59,9 | 6,26 | 4,13 | 3,57 |
| **78** | normal | 36,5 | 5,16 | 3,49 | 2,35 |
| **79** | normal | 26 | 4,66 | 4,37 | 3,3 |
| **80** | normal | 46,9 | 5,71 | 4,84 | 3,6 |
| **81** | normal | 42,2 | 5,83 | 3,91 | 2,7 |
| **82** | normal | 12 | 2,48 | 2,77 | 2,52 |
| **83** | normal | 21,6 | 6,55 | 3,46 | 2,5 |
| **84** | normal | 32,6 | 4,71 | 4,27 | 3,96 |
| **85** | normal | 57,4 | 5,2 | 4,13 | 2,38 |
| **86** | normal | 71,5 | 6,87 | 5,52 | 2,94 |
| **87** | normal | 103 | 8,12 | 6,09 | 2,58 |
| **88** | normal | 42,5 | 4,65 | 3,42 | 2,34 |
| **89** | normal | 23,1 | 4,46 | 3,3 | 2,31 |
| **90** | normal | 73,5 | 6,67 | 4,52 | 2,71 |
| **91** | normal | 92,8 | 8,47 | 5,2 | 3,2 |
| **92** | normal | 120,2 | 10,2 | 5,21 | 3,76 |
| Average | | 57,41 | 6,46 | 4,45 | 2,83 |

seedl hgt = seedling height; diam at b/h = diameter at breast height, diam at mid = diameter at middle point of stem height, diam at tip = diameter at the the tip of the stem height

**Supplementary Table 3.2. List of consensus and allele-specific primers of the apple S-RNase.**

| ***S*-allele** | **Primer name** | **Sequence 5′–3′** | **PCR product (bp)** |
|---|---|---|---|
| SRB | [a]Consensus | GTT CAY GGD TTR TGC CTT C<br>GGC CAA ATH ATD DYC ARC YG | multiple |
| $S_1$ | [c]MdS1SpF<br>[c]MdS1SpR | TGT AAG GCA CCG CCA TAT CAT AC<br>CAA CCT CAA CCA ATT CAG TCA ATG A | 734 |
| $S_2$ | [c]MdS2SpF<br>[c]MdS2SpR | AAC ATGAATCGAAGTGAATTATTTA<br>TTG AGG TTT GGT TCC TTA CCA TG | 489 |
| $S_3$ | [c]MdS3SpF<br>[c]MdS3SpR | GGC GAA AAT TAA ACC GGA GAA GAA<br>CCT CTC GTC CTA TAT ATG GAA ATC AC | 292 |
| $S_5$ | [c]MdS5SpF<br>[c]MdS5SpR | GGT CAA ACC CAC GGC GTC TCA<br>ATT CAG TTA TCC CAT TCT TCG | 1447 |
| $S_7$ | [c]MdS7SpF<br>[c]MdS7SpR | AGTAAATCAACCGTGGATGCTCAG<br>TTACAATATCTACCTGTTTCCTGGG | 397 |
| $S_9$ | [c]MdS9SpF<br>[c]MdS9SpR | CCACTTTAATCCTACTCCTTGTAGA<br>TCAATTTCCTTCTGTGTCCTGAATT | 522 |
| $S_{10}$ | [b]FTC12<br>[c]MdS10SpR | CCA AAC GTA CTC AAT CGA AG<br>TCC CGT GTC CTG AAT CTC CC | 203 |
| $S_{24}$ | [c]MdS24SpF<br>[c]MdS24SpR | ATGGCTCCTGTGCGTCTTCCC<br>CGTCATCCGTGTATAGGGCAACT | 421 |

[a] Bošković (unpublished), [b]Broothaerts 2003, [c]Long *et al*., 2010

**Supplementary Table 4.1 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' genetic linkage map.**

| Locus | Position (cM) | Parental Segregation | Observed Segregation | Expected segregation | $\chi^2$ | df | Significance |
|---|---|---|---|---|---|---|---|
| SNP_GDsnp02575 | 48.724 | <lmxll> | 43:49 | 1:1 | 0.39 | 1 | |
| SNP_FB_1034394 | 49.835 | <lmxll> | 44:47 | 1:1 | 0.10 | 1 | |
| SNP_FB_0762865 | 49.835 | <lmxll> | 44:48 | 1:1 | 0.17 | 1 | |
| *crinkledw2* | 49.835 | <hkxhk> | 68:24 | 3:1 | 33.57 | 1 | ******** |
| SNP_FB_1103570 | 50,965 | <lmxll> | 43:49 | 1:1 | 0.39 | 1 | |
| SNP_FB_1035415 | 52,096 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_RB_31145617 | 52,096 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_FB_0764489 | 52,096 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_FB_0764491 | 52,096 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_FB_0764512 | 52,096 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_FB_1035415 | 53,227 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_RB_31145617 | 53,227 | <lmxll> | 42:50 | 1:1 | 0.70 | 1 | |
| SNP_GDsnp00975 | 53,227 | <lmxll> | 41:50 | 1:1 | 0.89 | 1 | |
| SNP_FB_0765104 | 54,357 | <lmxll> | 41:51 | 1:1 | 1.09 | 1 | |
| SNP_FB_0765111 | 54,357 | <lmxll> | 41:51 | 1:1 | 1.09 | 1 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *crinkledw1* | 54,357 | <hkxhk> | 24:68 | 1:3 | 33.57 | 1 | ******** |
| SNP_FB_1053497 | 54,357 | <lmxll> | 52:40 | 1:1 | 1.57 | 1 | |
| SNP_FB_1053497 | 54,357 | <lmxll> | 40:50 | 1:1 | 1.11 | 1 | |
| SNP_FB_0765586 | 54,357 | <lmxll> | 42:52 | 1:1 | 1.57 | 1 | |
| SNP_RB_32814468 | 54,357 | <lmxll> | 43:52 | 1:1 | 1.57 | 1 | |
| SNP_FB_0461117 | 54,357 | <lmxll> | 44:52 | 1:1 | 1.57 | 1 | |
| SNP_FB_0461119 | 57,768 | <lmxll> | 44:52 | 1:1 | 1.57 | 1 | |
| SNP_FB_0765594 | 57,768 | <lmxll> | 44:52 | 1:1 | 1.57 | 1 | |
| SNP_FB_0765947 | 57,768 | <lmxll> | 46:52 | 1:1 | 1.57 | 1 | |

**Supplementary Table 4.2 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 2 of the 'M.1' genetic linkage map.**

| **Locus** | **Position (cM)** | **Parental Segregation** | **Observed Segregation** | **Expected segregation** | **$\chi^2$** | **df** | **Significance** |
|---|---|---|---|---|---|---|---|
| FB_0451368 | 4,522 | <nnxnp> | 58:32 | 1:1 | 6.26 | 1 | ** |
| *crinkledw1* | 5,633 | <hkxhk> | 68:24 | 3:1 | 33.57 | 1 | ******** |
| FB_0949262 | 5,633 | <nnxnp> | 57:35 | 1:1 | 5.26 | 1 | ** |
| *crinkledw2* | 5,633 | <hkxhk> | 24:68 | 3:1 | 33.57 | 1 | ******** |
| FB_0452379 | 6,744 | <nnxnp> | 58:34 | 1:1 | 6.26 | 1 | ** |

**Supplementary Table 4.3 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' genetic linkage map.**

| Locus | Position (cM) | Parental Segregation | Observed Segregation | Expected segregation | $\chi^2$ | df | Significance |
|---|---|---|---|---|---|---|---|
| SNP_FB_1033686 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_1033701 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_1033703 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_RB_28804875 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759430 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759435 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759441 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759442 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| crinkledw1 | 232.246 | <hkxhk> | 23:62 | 1:3 | 0.19 | 2 | |
| GDsnp02575 | 232.246 | <lmxll> | 42:43 | 1:1 | 0.01 | 1 | |
| crinkledw2 | 232.246 | <hkxhk> | 62:23 | 3:1 | 30.34 | 2 | ******* |
| SNP_FB_0811041 | 272.413 | <lmxll> | 40:45 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811048 | 272.413 | <lmxll> | 40:45 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811059 | 272.413 | <lmxll> | 45:40 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811056 | 272.413 | <hkxhk> | 11:48:26 | 1:2:1 | 6.72 | 2 | ** |
| SNP_FB_0767173 | 273.603 | <hkxhk> | 25:48:12 | 1:2:1 | 5.4 | 2 | * |
| SNP_GDsnp00246 | 273.603 | <lmxll> | 39:46 | 1:1 | 0.58 | 1 | |

**Supplementary Table 4.4 Segregation patterns for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'McIntosh' consensus genetic linkage map.**

| Locus | Position (cM) | Parental Segregation |
|---|---|---|
| SNP_FB_0971471 | 270.420 | <llxlm> |
| SNP_FB_1033686 | 271.600 | <llxlm> |
| SNP_FB_1033701 | 271.600 | <llxlm> |
| SNP_FB_1033703 | 271.600 | <llxlm> |
| SNP_RB_28804875 | 271.600 | <llxlm> |
| SNP_FB_0759430 | 271.600 | <llxlm> |
| SNP_FB_0759435 | 271.600 | <llxlm> |
| SNP_FB_0759441 | 271.600 | <llxlm> |
| SNP_FB_0759442 | 271.600 | <llxlm> |
| *crinkledw1* | 272.801 | <hkxhk> |
| GDsnp02575 | 272.801 | <hkxhk> |
| *crinkledw2* | 272.801 | <hkxhk> |
| SNP_FB_0811041 | 283.228 | <llxlm> |
| SNP_FB_0811048 | 283.228 | <llxlm> |
| SNP_FB_0811059 | 283.228 | <llxlm> |
| SNP_FB_0811056 | 283.228 | <hkxhk> |
| SNP_FB_0767173 | 284.418 | <hkxhk> |
| GDsnp00246 | 284.418 | <hkxhk> |

**Supplementary Table 4.5 Monogenic segregation values for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 2 of the 'M.1' genetic linkage map.**

| Locus | Position (cM) | Parental Segregation | Observed Segregation | Expected segregation | $\chi^2$ | df | Significance |
|---|---|---|---|---|---|---|---|
| SNP_FB_1033686 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_1033701 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_1033703 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_RB_28804875 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759430 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759435 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759441 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| SNP_FB_0759442 | 203.583 | <lmxll> | 43:42 | 1:1 | 0.01 | 1 | |
| *crinkledw1* | 232.246 | <hkxhk> | 23:62 | 1:3 | 0.19 | 2 | |
| GDsnp02575 | 232.246 | <lmxll> | 42:43 | 1:1 | 0.01 | 1 | |
| *crinkledw2* | 232.246 | <hkxhk> | 62:23 | 3:1 | 30.34 | 2 | ******* |
| SNP_FB_0811041 | 272.413 | <lmxll> | 40:45 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811048 | 272.413 | <lmxll> | 40:45 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811059 | 272.413 | <lmxll> | 45:40 | 1:1 | 0.29 | 1 | |
| SNP_FB_0811056 | 272.413 | <hkxhk> | 11:48:26 | 1:2:1 | 6.72 | 2 | ** |
| SNP_FB_0767173 | 273.603 | <hkxhk> | 25:48:12 | 1:2:1 | 5.4 | 2 | * |
| GDsnp00246 | 273.603 | <lmxll> | 39:46 | 1:1 | 0.58 | 1 | |

**Supplementary Table 4.6 Segregation patterns for the crinkle dwarf locus and its flanking single nucleotide polymorphism (SNP) markers in linkage group (LG) 8 of the 'M.1' consensus genetic linkage map.**

| Locus | Position (cM) | Parental Segregation |
|---|---|---|
| SNP_FB_0758077 | 75.759 | <hkxhk> |
| SNP_FB_0758061 | 75.759 | <hkxhk> |
| SNP_FB_0758063 | 75.759 | <hkxhk> |
| SNP_FB_0758068 | 75.759 | <hkxhk> |
| SNP_FB_0891072 | 79.419 | <hkxhk> |
| SNP_RB_28185282 | 79.419 | <hkxhk> |
| *crinkledw1* | 191.691 | <hkxhk> |
| *crinkledw2* | 191.691 | <hkxhk> |
| SNP_FB_0811056 | 261.597 | <hkxhk> |
| SNP_FB_0767173 | 262.788 | <hkxhk> |
| SNP_FB_0754071 | 292.413 | <hkxhk> |
| SNP_FB_0132994 | 292.413 | <nnxnp> |
| SNP_FB_0753434 | 292.413 | <nnxnp> |
| SNP_FB_0753471 | 292.413 | <nnxnp> |
| SNP_FB_0755022 | 292.413 | <nnxnp> |
| SNP_FB_0755026 | 292.413 | <nnxnp> |
| SNP_FB_0132990 | 292.413 | <nnxnp> |
| SNP_FB_0932949 | 292.413 | <nnxnp> |

**Supplementary Table 5.1 Raw, clean and mapped reads between each Pools of normal and crinkle dwarf phenotypes.**

| Sample ID | Replicate | Paired raw reads | Paired cleaned reads | Paired mapped reads | Properly paired mapped reads | Properly paired transcript |
|---|---|---|---|---|---|---|
| **Pool1** | L001 | 55 833 448 | 38 133 220 | 33 222 324 | 28 498 516 | 58 472 |
| **Pool1** | L002 | 56 058 600 | 38 335 562 | 33 388 382 | 28 650 362 | 58 587 |
| **Pool2** | L001 | 57 694 520 | 41 637 508 | 36 808 822 | 32 259 326 | 59 468 |
| **Pool2** | L002 | 58 030 912 | 41 930 696 | 37 066 430 | 32 499 990 | 59 536 |
| **Pool3** | L001 | 59 088 338 | 42 566 700 | 37 433 332 | 32 684 188 | 60 069 |
| **Pool3** | L002 | 59 331 224 | 42 790 970 | 37 622 643 | 32 862 656 | 60 198 |
| **Pool4** | L001 | 53 243 956 | 38 885 320 | 34 24 8867 | 29 738 334 | 61 027 |
| **Pool4** | L002 | 53 559 498 | 39 160 080 | 34 48 9135 | 29 962 994 | 61 123 |
| **Pool5** | L001 | 53 460 502 | 37 955 778 | 33 04 7046 | 28 914 280 | 59 741 |
| **Pool5** | L002 | 53 713 442 | 38 175 928 | 33 233 578 | 29 083 854 | 59 745 |
| **Pool6** | L001 | 46 928 508 | 33 180 844 | 29 264 399 | 25 467 292 | 60 475 |
| **Pool6** | L002 | 47 230 880 | 33 433 266 | 29 486 244 | 25 673 776 | 60 504 |
| **Total** | | **654 173 828** | **466 185 872** | **409 311 202** | **356 295 568** | **718 945** |

**Supplementary Table 5.2 Quality verification of the RNA-seq library synthesized from the pooled samples between normal and crinkle dwarf phenotypes.**

| Sample ID | Index | Total RNA start conc. (ng/ µL) | fragmentation time (min) | No. of PCR cycles | cDNA conc. (ng/ µL) |
|---|---|---|---|---|---|
| **Pool1** | AR001 | ±1000 | 8 | 12 | 22.0 |
| **Pool2** | AR002 | ±1000 | 8 | 12 | 47.7 |
| **Pool3** | AR003 | ±1000 | 8 | 12 | 40.6 |
| **Pool4** | AR004 | ±1000 | 8 | 12 | 49.7 |
| **Pool5** | AR005 | ±1000 | 8 | 12 | 57.0 |
| **Pool6** | AR006 | ±1000 | 8 | 12 | 56.0 |

conc.= concentration

**Supplementary Table 5.3 COG functional classification of up- and down-regulated genes between normal and crinkle dwarf phenotypes.**

| Function Class | Code | Up-regulated | Down-regulated |
|---|---|---|---|
| **Information Storage and Processing** | | | |
| Transcription | K | 58 | 9 |
| Translation, ribosomal structure and biogenesis | J | 5 | 1 |
| Replication, recombination and repair | L | 5 | 4 |
| RNA processing and modification | A | 4 | 1 |
| Chromatin structure and dynamics | B | 0 | 0 |
| **Cellular Processes and Signaling** | | | |
| Signal transduction mechanisms | T | 144 | 4 |
| Posttranslational modification, protein turnover, chaperones | O | 58 | 2 |
| Intracellular trafficking, secretion, and vesicular transport | U | 6 | 1 |
| Cell wall/membrane/envelope biogenesis | M | 6 | 0 |
| Defense mechanisms | V | 3 | 1 |
| Cytoskeleton | Z | 2 | 0 |
| Cell cycle control, cell division, chromosome partitioning | D | 1 | 0 |
| Nuclear structure | Y | 0 | 0 |
| Extracellular structures | W | 0 | 0 |
| Cell motility | N | 0 | 0 |
| **Metabolism** | | | |
| Secary metabolites biosynthesis, transport and catabolism | Q | 57 | 0 |
| Carbohydrate transport and metabolism | G | 29 | 0 |
| Inorganic ion transport and metabolism | P | 20 | 2 |
| Amino acid transport and metabolism | E | 19 | 2 |
| Lipid transport and metabolism | I | 15 | 7 |
| Energy production and conversion | C | 11 | 3 |
| Coenzyme transport and metabolism | H | 4 | 1 |
| Nucleotide transport and metabolism | F | 3 | 0 |
| **Poorly Characterised** | | | |
| Function unknown | S | 207 | 32 |
| General function prediction only | R | 0 | 0 |

**Supplementary Table 5.4 Enzyme mapping and annotations according to KEGG for up-regulated DEGs.**

| Enzyme Name | Pathway Name | Pathway ID | #enzymes in Pathway | #Seqs in Pathway | Seqs |
|---|---|---|---|---|---|
| ec:4.2.1.1 - anhydrase | Nitrogen metabolism | map00910 | 2 | 1 | XM_008387117.3 |
| ec:2.4.1.69 - 1 galactoside alpha-(1,2)-fucosyltransferase | Glycosphingolipid biosynthesis - lacto and neolacto series, Glycosphingolipid biosynthesis - globo and isoglobo series | map00601, map00603 | 1, 1 | 1 | XM_008341646.3 |
| ec:2.3.1.5 - N-acetyltransferase | Drug metabolism - other enzymes, Caffeine metabolism, Nitrotoluene degradation | map00983, map00232, map00633 | 3, 1, 1 | 1 | XM_008347481.3 |
| ec:2.4.1.123 - 3-alpha-galactosyltransferase | Galactose metabolism | map00052 | 1 | 2 | XM_008344971.3, XM_008372139.3 |
| ec:1.2.1.41 - dehydrogenase | Biosynthesis of antibiotics, Carbapenem biosynthesis, Arginine and proline metabolism | map01130, map00332, map00330 | 3, 1, 1 | 1 | XM_008389606.3 |
| ec:3.2.1.1 - glycogenase | Starch and sucrose metabolism | map00500 | 1 | 2 | XM_008374925.3, XM_008348916.3 |
| ec:1.10.3.1 - oxidase | Isoquinoline alkaloid biosynthesis, Tyrosine metabolism | map00950, map00350 | 1, 1 | 1 | NM_001319261.1 |
| ec:3.4.19.13 - gamma-glutamate hydrolase | Glutathione metabolism | map00480 | 2 | 1 | XM_008388808.3 |
| ec:1.1.1.27 - dehydrogenase | Biosynthesis of antibiotics, Propanoate metabolism, Glycolysis / Gluconeogenesis, Pyruvate metabolism, Cysteine and methionine metabolism | map01130, map00640, map00010, map00620, map00270 | 3, 1, 1, 2, 1 | 1 | XM_008342322.3 |
| ec:1.11.1.7 - | Phenylpropanoid biosynthesis | map00940 | 1 | 5 | XM_008390100.3, XM_008386795.3, |

| | | | | | |
|---|---|---|---|---|---|
| lactoperoxidase | | | | | XM_008383760.3, XM_008371099.3, XM_008386507.3 |
| ec:3.6.1.3 - adenylpyrophosphatase | Purine metabolism | map00230 | 2 | 4 | XM_008352189.3, XM_008391708.3, XM_008351564.3, XM_008371602.3 |
| ec:3.1.1.32 - A1 | alpha-Linolenic acid metabolism, Glycerophospholipid metabolism | map00592, map00564 | 1, 1 | 1 | XM_008364303.3 |
| ec:3.1.3.2 - phosphatase | Thiamine metabolism, Riboflavin metabolism | map00730, map00740 | 2, 1 | 4 | XM_008372059.3, XM_008372060.3, XM_008376800.3, XM_008372061.3 |
| ec:1.14.13.8 - monooxygenase | Drug metabolism - cytochrome P450 | map00982 | 2 | 1 | XM_008339660.3 |
| ec:3.1.1.1 - ali-esterase | Drug metabolism - other enzymes | map00983 | 3 | 3 | XM_008365497.3, XM_008364303.3, XM_008349560.3 |
| ec:3.6.1.15 - phosphatase | Purine metabolism, Thiamine metabolism | map00230, map00730 | 2, 2 | 4 | XM_008352189.3, XM_008391708.3, XM_008351564.3, XM_008371602.3 |
| ec:3.1.1.11 - pectin demethoxylase | Pentose and glucuronate interconversions | map00040 | 1 | 2 | XM_008365497.3, XM_008349560.3 |
| ec:5.4.99.7 - synthase | Steroid biosynthesis, Biosynthesis of antibiotics | map00100, map01130 | 1, 3 | 1 | XM_017330640.2 |
| ec:4.4.1.5 - lyase | Pyruvate metabolism | map00620 | 2 | 1 | XM_017330618.2 |
| ec:3.2.1.14 - ChiC | Amino sugar and nucleotide sugar metabolism | map00520 | 1 | 1 | XM_008370975.3 |
| ec:3.1.3.16 - phosphatase | Th1 and Th2 cell differentiation, PD-L1 expression and PD-1 checkpoint pathway in cancer, T cell receptor signaling pathway | map04658, map05235, map04660 | 1, 1, 1 | 1 | XM_008370527.3 |
| ec:6.3.1.2 - synthetase | Alanine, aspartate and glutamate metabolism, Arginine biosynthesis, | map00250, map00220, | 1, 1, 2, 1 | 1 | XM_008342223.2 |

| | | | | | |
|---|---|---|---|---|---|
| | Nitrogen metabolism, Glyoxylate and dicarboxylate metabolism | map00910, map00630 | | | |
| ec:2.5.1.18 - transferase | Metabolism of xenobiotics by cytochrome P450, Glutathione metabolism, Drug metabolism - other enzymes, Drug metabolism - cytochrome P450 | map00980, map00480, map00983, map00982 | 1, 2, 3, 2 | 1 | XM_008366006.3 |

**Supplementary Table 5.5 Enzyme mapping and annotations according to KEGG for down-regulated DEGs.**

| Enzyme Name | Pathway Name | Pathway ID | #enzymes in pathway | #Seqs in Pathway | Seqs |
|---|---|---|---|---|---|
| ec:4.2.1.1 - anhydrase | Nitrogen metabolism | map00910 | 1 | 1 | gi\|1039832438\|ref\|XM_008385755.2\| |
| ec:2.1.1.4 - O-methyltransferase | Tryptophan metabolism | map00380 | 1 | 1 | gi\|1039885915\|ref\|XM_008378603.2\| |
| ec:2.7.1.25 - kinase | Purine metabolism, Sulfur metabolism | map00230, map00920 | 3, 2 | 1 | gi\|1039847902\|ref\|XM_008340218.2\| |
| ec:4.4.1.9 - synthase | Cyanoamino acid metabolism | map00460 | 1 | 1 | gi\|1039895905\|ref\|XM_008345184.2\| |
| ec:1.14.17.4 - oxidase | Cysteine and methionine metabolism | map00270 | 6 | 1 | gi\|1039848394\|ref\|XM_008339460.2\| |
| ec:2.3.1.188 - O-feruloyl transferase | Cutin, suberine and wax biosynthesis | map00073 | 1 | 1 | gi\|1039925767\|ref\|XM_008363355.2\| |
| ec:1.1.1.95 - dehydrogenase | Methane metabolism, Cysteine and methionine metabolism, Biosynthesis of antibiotics, Glycine, serine and threonine metabolism | map00680, map00270, map01130, map00260 | 1, 6, 3, 1 | 1 | gi\|1039892639\|ref\|XM_008382040.2\| |
| ec:3.6.1.3 - adenylpyrophosphatase | Purine metabolism | map00230 | 3 | 1 | gi\|1039928363\|ref\|XR_528258.2\| |
| ec:3.1.3.2 - phosphatase | Thiamine metabolism, Riboflavin metabolism | map00730, map00740 | 2, 1 | 1 | gi\|1039921058\|ref\|XM_008359862.2\| |
| ec:2.1.1.42 - 3'-O-methyltransferase | Flavone and flavonol biosynthesis | map00944 | 2 | 1 | gi\|1039885915\|ref\|XM_008378603.2\| |
| ec:1.14.11.13 - 2beta-dioxygenase | Diterpenoid biosynthesis | map00904 | 1 | 1 | gi\|1039831208\|ref\|XM_008385804.2\| |
| ec:3.6.1.15 - | Thiamine metabolism, Purine | map00730, | 2, 3 | 1 | gi\|1039879099\|ref\|XR_529390.2\| |

| | | | | | |
|---|---|---|---|---|---|
| phosphatase | metabolism | map00230 | | | |
| ec:4.4.1.14 - synthase | Cysteine and methionine metabolism | map00270 | 6 | 1 | gi\|1039868682\|ref\|XM_008344016.2\| |
| ec:2.1.1.76 - 3-O-methyltransferase | Flavone and flavonol biosynthesis | map00944 | 2 | 1 | gi\|1039885915\|ref\|XM_008378603.2\| |
| ec:2.5.1.29 - diphosphate synthase | Terpenoid backbone biosynthesis, Biosynthesis of antibiotics | map00900, map01130 | 1, 3 | 1 | gi\|1039893737\|ref\|XM_008383282.2\| |
| ec:1.14.19.2 - 9-desaturase | Fatty acid biosynthesis, Biosynthesis of unsaturated fatty acids | map00061, map01040 | 1, 1 | 1 | gi\|1039899843\|ref\|XM_008347059.2\| |
| ec:2.5.1.47 - synthase | Sulfur metabolism, Cysteine and methionine metabolism, Biosynthesis of antibiotics | map00920, map00270, map01130 | 2, 6, 3 | 1 | gi\|1039895905\|ref\|XM_008345184.2\| |
| ec:4.4.1.11 - gamma-lyase | Cysteine and methionine metabolism, Selenocompound metabolism | map00270, map00450 | 6, 1 | 1 | gi\|658060383\|ref\|XM_008367807.1\| |
| ec:2.5.1.6 - adenosyltransferase | Cysteine and methionine metabolism | map00270 | 6 | 1 | gi\|1039873073\|ref\|XM_008370701.2\| |

## Supplementary Figure 5.1

Workflow and stepwise analysis used in transcriptome profiling of the pools between normal *versus* crinkle dwarf phenotypes.